DATA WAREHOUSING AND DATA MINING LAB (R20A1283)

LAB MANUAL

B.TECH (III YEAR – II SEM) (2023-24)



Prepared By: Mrs.T.Shilpa Mr. P.Harikrishna Mr. P.V.Naresh

DEPARTMENT OF INFORMATION TECHNOLOGY

MALLA REDDY COLLEGE OF ENGINEERING&TECHNOLOGY (Autonomous Institution – UGC, Govt. of India)

Recognized under 2(f) and 12 (B) of UGC ACT 1956

Affiliated to JNTUH, Hyderabad, Approved by AICTE - Accredited by NBA & NAAC – 'A' Grade - ISO 9001:2008 Certified) Maisammaguda, Dhulapally (Post Via. Hakimpet), Secunderabad – 500100, Telangana State, India

DEPARTMENT OF INFORMATION TECHNOLOGY

VISION

To achieve high quality education in technical education that provides the skills and attitude to adapt to the global needs of the Information technologysector, through academic and research excellence.

MISSION

- To equip the students with the cognizance for problem solving and to improve the teaching learning pedagogy by using innovative techniques.
- To strengthen the knowledge base of the faculty and students with the motivation towards possession of effective academic skills and relevant research experience.
- To promote the necessary moral and ethical values among the engineers for the betterment of the society.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO1 – PROFESSIONALISM & CITIZENSHIP

To create and sustain a community of learning in which students acquire knowledge and learn to apply it professionally with due consideration for ethical, ecological and economic issues.

PEO2 – TECHNICAL ACCOMPLISHMENTS

To provide knowledge based services to satisfy the needs of society and the industry by providing hands on experience in various technologies in core field.

PEO3 – INVENTION, INNOVATION AND CREATIVITY

To make the students to design, experiment, analyze, interpret in the core field with the help of other multi-disciplinary concepts wherever applicable.

PEO4 – PROFESSIONAL ETHICS

To educate the students to disseminate research findings with good soft skills and become a successful entrepreneur.

PEO5 – HUMAN RESOURCE DEVELOPMENT

To graduate the students in building national capabilities in technology, education and research.

PROGRAM SPECIFIC OUTCOMES (PSOs)

After the completion of the course, B. Tech Information Technology, the graduates will have the following Program Specific Outcomes:

- 1. **Fundamentals and critical knowledge of the Computer System:-** Able to Understand the working principles of the computer System and its components, Apply the knowledge to build, asses, and analyze the software and hardware aspects of it.
- 2. The comprehensive and Applicative knowledge of Software Development: Comprehensive skills of Programming Languages, Software process models, methodologies, and able to plan, develop, test, analyze, and manage the software and hardware intensive systems in heterogeneous platforms individually or working in teams.
- 3. Applications of Computing Domain & Research: Able to use the professional, managerial, interdisciplinary skill set, and domain specific tools in development processes, identify the research gaps, and provide innovative solutions to them.

PROGRAM OUTCOMES (POs)

Engineering Graduates will be able to:

- 1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- 2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- 3. **Design / development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- 5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
- 6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- 7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- 8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- 9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
- 10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
- 11. **Project management and finance** : Demonstrate knowledge and understandingof the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multi disciplinary environments.
- 12. Life- long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.



MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

Maisammaguda, Dhulapally Post, Via Hakimpet, Secunderabad – 500100

DEPARTMENT OF INFORMATION TECHNOLOGY GENERAL LABORATORY INSTRUCTIONS

- 1. Students are advised to come to the laboratory at least 5 minutes before (to the starting time), those who come after 5 minutes will not be allowed into the lab.
- 2. Plan your task properly much before to the commencement, come prepared to the lab with the synopsis / program / experiment details.
- 3. Student should enter into the laboratory with:
- a. Laboratory observation notes with all the details (Problem statement, Aim, Algorithm, Procedure, Program, Expected Output, etc.,) filled in for the lab session.
- b. Laboratory Record updated up to the last session experiments and other utensils (if any) needed in the lab.
- c. Proper Dress code and Identity card.
- 4. Sign in the laboratory login register, write the TIME-IN, and occupy the computer system allotted to you bythe faculty.
- 5. Execute your task in the laboratory, and record the results / output in the lab observation note book, and get certified by the concerned faculty.
- 6. All the students should be polite and cooperative with the laboratory staff, must maintain the discipline and decency in the laboratory.
- 7. Computer labs are established with sophisticated and high-end branded systems, which should be utilized properly.
- 8. Students / Faculty must keep their mobile phones in SWITCHED OFF mode during the lab sessions. Misuse of the equipment, misbehaviors with the staff and systems etc., will attract severe punishment.
- 9. Students must take the permission of the faculty in case of any urgency to go out; if anybody found loitering outside the lab / class without permission during working hours will be treated seriously and punished appropriately.
- 10. Students should LOG OFF/ SHUT DOWN the computer system before he/she leaves the lab after completing the task (experiment) in all aspects. He/she must ensure the system / seat is kept properly.

HEAD OF THE DEPARTMENT

PRINCIPAL

MALLA REDDY COLLEGE OF ENGINEERING AND TECHNOLOGYB.TECH - III- YEAR II-SEM-ITL/T/P/C

-/-/3/1.5

(R20A1283)DATA WAREHOUSING AND DATA MINING

COURSE OBJECTIVES:

- 1. To identify how to build a data warehouse and query it (using open source tools like Pentaho Data Integration and Pentaho Business Analytics)
- 2. To get an understanding of data mining tasks using a data mining toolkit (such as open source WEKA)
- 3. To understand the data sets and data preprocessing
- 4. To study the working of algorithms for data mining tasks such as association rulemining, classification, clustering and regression
- 5. To get acquainted to the data mining techniques with varied input values for different parameters

LIST OF EXPERIMENTS: Experiments using Weka / Clementine Tools

- 1. Installation of WEKA Tool
- 2. Creating new Arff File
- 3. Data Processing Techniques on Data set.
- 4. Data cube construction OLAP operations
- 5. Implementation of Apriori algorithm
- 6. Implementation of FP Growth algorithm
- 7. Implementation of Decision Tree Induction
- 8. Calculating Information gain measures
- 9. Classification of data using Bayesian approach
- 10. Classification of data using K Nearest Neighbor Approach.
- 11.Implementation of K means algorithm

COURSE OUTCOMES:

- 1. To specify which data processing technique can be applied
- 2. To construct a Data cube
- 3. To implement mining algorithms as a component to the existing tools
- 4. To differentiate the approaches based on classification
- 5. To associate mining techniques for realistic data

DATA WAREHOUSING AND DATA MINING LAB

Table of Contents

S.No	Name of the Experiment	Page No.
1	Installation of WEKA Tool	1
2	Creating new Arff File	11
3	Data Processing Techniques on Dataset	16
4	Data cube construction – OLAPoperations	25
5	Implementation of Apriori algorithm	33
6	Implementation of FP- Growthalgorithm	45
7	Implementation of Decision TreeInduction	54
8	Calculating Information gainsmeasures	63
9	Classification of data using Bayesianapproach	68
10	Implementation of K-meansalgorithm	76
11	Case Study: Create Placement.arff file to identify the students who are eligible for placements using KNN	85

DATAWARE HOUSE TOOLS

Cloudera	cloudera
Teradata	TERADATA
Oracle	ORACLE
TabLeau	+++++++ a b e a u *
ODEN S	OURCE DATA MINING TOOLS

OPEN SOURCE DATA MINING TOOLS

WEKA	Te University Waikato
Orange	orange DATA MINING FRUITFUL&FUN
KNIME	
R-Programming	

Experiment 1: Installation of WEKA Tool

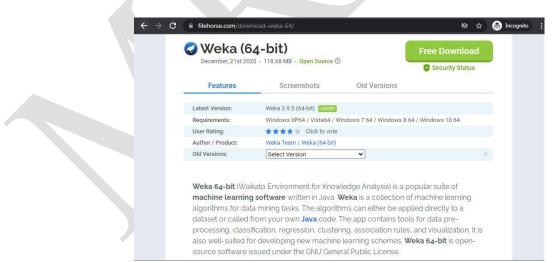
Introduction

Weka (pronounced to rhyme with Mecca) is a workbench that contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. The original non-Java version of Weka was a Tcl/Tk front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C, and Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research. Advantages of Weka include:

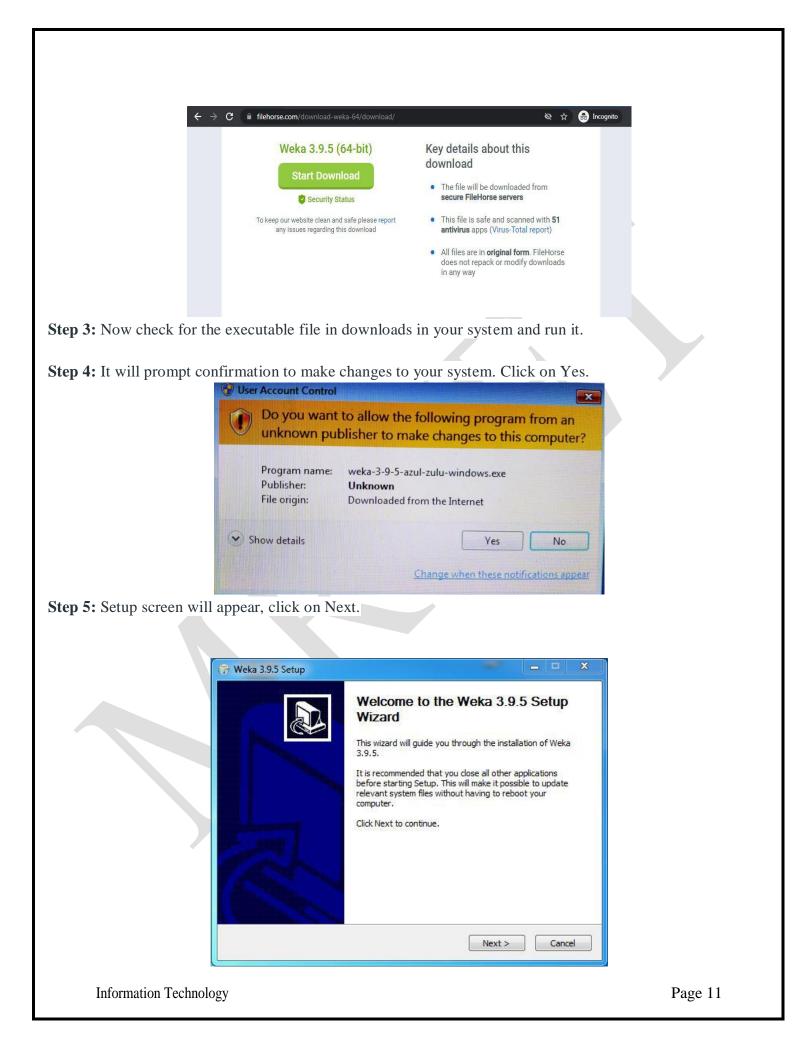
- Free availability under the GNU General Public License.
- Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform
- A comprehensive collection of data preprocessing and modeling techniques
- Ease of use due to its graphical user interfaces

Installing Weka on Windows:

Step 1: Visit this <u>website</u> using any web browser. Click on Free Download.



Step 2: It will redirect to a new webpage, click on Start Download. Downloading of the executable file will start shortly. It is a big 118 MB file that will take some minutes



Step 6: The next screen will be of License Agreement, click on I Agree.

Weko	License Agre					
	Please review	the license term	is before install	ing Weka	a 3.9.5.	
Press Page Down to see	the rest of the agr	eement.				
and the second	RAL PUBLIC LICENS	E				*
version 3	, 29 June 2007				1	
Copyright (C) 2007 Fre Everyone is permitted t of this license documen	to copy and distribu	te verbatim copi				
Pream	ble					
The GNU General Publi software and other kind		copylef <mark>t lic</mark> ense	for			Ŧ
If you accept the terms agreement to install Wek		dick I Agree to c	ontinue. You m	ust acce	pt the	
soft Install System v08-	Mar-2013.cvs ——					

Step 7: Next screen is of choosing components, all components are already marked so don't change anything just click on the Install button.

Weka	Choose Components Choose which features of We	ka 3.9.5 you want <mark>to</mark> install.
nstall. Click Next to continue		components you don't want to
Select the type of install:	Full	
Or, select the optional components you wish to install:	Associate Files	Description Position your mouse over a component to see its description.
Space required: 301.0MB		
soft Install System v08-Mar-	2013.cvs	

Step 8: The next screen will be of installing location so choose the drive which will have sufficient memory space for installation. It needed a memory space of 301 MB.

Weka	Choose Install Location Choose the folder in which to install Weka 3.9.5.	
	.9.5 in the following folder. To install in a different folder, or, Click Next to continue.	dick Browse
Destination Folder C:\Program Files\W	eka-3-9-5 Brows	2
C:\Program Files\W Space required: 301.0M	IB	5
	IB 3	B

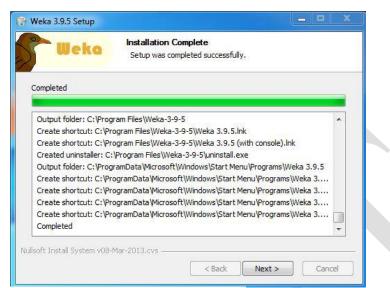
Step 9: Next screen will be of choosing the Start menu folder so don't do anything just click on Install Button.

Weka 3.9.5 Setup			L.	>
Weka		art Menu Folder tart Menu folder fo	r the Weka 3.9.5 sh	ortcuts,
Select the Start Menu fo can also enter a name to Weka 3.9.5			e the program's shor	tcuts. You
Accessories Administrative Tools Adobe AnyDesk Flex Windows Free Cam 8 Games Intel PROSet Wireless Inter Download Mar IP Messenger for Win Java	iager			A III
Do not create shortc lsoft Install System v08-	7.77	< Back	Install	Cancel

Step 10: After this installation process will start and will hardly take a minute to complete the installation.

Please wait while Weka 3.9.5 is being installer	
Extract: MultiScheme.html 100% Extract: RandomCommittee.html 100% Extract: RandomSubSpace.html 100% Extract: RandomizableFilteredClassifier.html 100% Extract: RegressionByDiscretization.html 100% Extract: Stacking.html 100% Extract: Vote.html 100% Extract: WeightedInstancesHandlerWrapper.html 100% Extract: weightedInstancesHandlerWrapper.html 100%	^

Step 11: Click on the Next button after the installation process is complete.



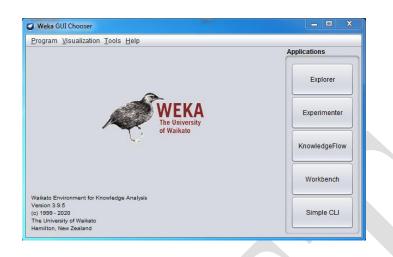
Step 12: Click on Finish to finish the installation process.

	Completing the Weka 3.9.5 Setup Wizard
	Weka 3.9.5 has been installed on your computer. Click Finish to close this wizard.
R	
	< Back Finish Cancel

Step 13: Weka is successfully installed on the system and an icon is created on the desktop.



Step 14: Run the software and see the interface.



WEKA Tool Description:

Open the program. Once the program has been loaded on the user's machine it is opened by navigating to the programs start option and that will depend on the users operating system. Figure 1.1 is an example of the initial opening screen on a computer.

There are four options available on this initial screen:

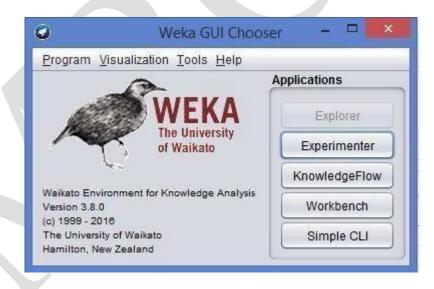


Fig: 1.1 Weka GUI

1. Explorer - the graphical interface used to conduct experimentation on raw data After clicking the Explorer button the weka explorer interface appears.

	0				
		Classify Cluster Ass	Weka Explorer		
	Attribute Evalua				
	Choose	CfsSubsetEval -P 1 -E 1			
	Search Method				
	Choose	0	weka.gui.GenericObjectEditor	×	
		weka.attributeSelection			
	Attribute Selec	About			
	 Use full ti Cross-va 	BestFirst:		More	
		Searches the space	ce of attribute subsets by greedy hillclimbin backtracking facility.		
	(Num) trend	1			
		direction	Forward		
	Start Result list (righ	lookupCacheSize	1		
	Result list (righ	-			
		searchTermination	5		
		startSet	r F		
	Status	Open	Save OK	Cancel	
	OK				Log
			Fig: 1.2 Pre-proces	ssor	Log
0			Fig: 1.2 Pre-proces	ssor	- • ×
Prep	rocess Classify	Cluster Associate		ssor	
Prep	rocess Classify te Evaluator	Cluster Associate	Weka Explorer		
Prep	te Evaluator	wek	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor	ssor	
Prep Attribu	te Evaluator 100se Ø Weka.a	wek ttributeSelection.CfsSub	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor		
Prep Attribu Cl	te Evaluator noose Method noose	wek ttributeSelection.CfsSub	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor		
Attribut Col Search	te Evaluator	wek ttributeSelection.CfsSub: : : : : : : :	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval	More	
Prep Attribu Col Search Col Attribu	te Evaluator noose weka.a About noose Cf: te Selec Ev lee full te	wek ttributeSelection.CfsSub ssubsetEval : aluates the worth of a su fuidual predictive ability of	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor vsetEval ubset of attributes by considering the of each feature along with the degree		- • ×
Attribu Search Attribu	te Evaluator noose weka.a About noose Cf: te Selec Ev lee full te	wek ttributeSelection.CfsSub : : sSubsetEval : aluates the worth of a su	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor vsetEval ubset of attributes by considering the of each feature along with the degree	More	
Attribu Search Attribu	te Evaluator 10058 @ 1 Method 10058 Cf: 10058 Cf:	wek ttributeSelection.CfsSub sSubsetEval : aluates the worth of a su fividual predictive ability of redundancy between the	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor setEval ubset of attributes by considering the of each feature along with the degree em	More Capabilities	
Attribu Search Attribu	te Evaluator 10058 @ 1 Method 10058 Cf: 10058 Cf:	wek ttributeSelection.CfsSub ssubsetEval : aluates the worth of a su fuidual predictive ability of	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor setEval ubset of attributes by considering the of each feature along with the degree em	More	- • ×
Attribu Searcl Cl Attribu	te Evaluator Noose Nethod Nethod Veka.a About Ct. Ev Jse full tr of	wek ttributeSelection.CfsSub- sSubsetEval : aluates the worth of a su fividual predictive ability of redundancy between the	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree am. False	More Capabilities	
Attribu Searcl Cl Attribu	te Evaluator	wek ttributeSelection.CfsSub s sSubsetEval : aluates the worth of a su dividual predictive ability of redundancy between the debug	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree am. False	More Capabilities	
Prep Attribu	te Evaluator Noose Method Verka.a About Toose Te Selec Cf: Ev Cf: Ev Cf: Cf: Cf: Cf: Cf: Cf: Cf: Cf:	wek ttributeSelection.CfsSub- t sSubsetEval : aluates the worth of a su fividual predictive ability redundancy between the debug oNotCheckCapabilities	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree am. False False True	More Capabilities	
Prep Attribu	te Evaluator Noose Method Veka.a About Noose te Selec Cf: Ev Inc of Cf: Ev Inc of d Start	wek ttributeSelection.CfsSub sSubsetEval : aluates the worth of a su dividual predictive ability of redundancy between the debug oNotCheckCapabilities locallyPredictive missingSeparate	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor ossetEval ubset of attributes by considering the of each feature along with the degree am. False False True False	More Capabilities	
Prep Attribu	te Evaluator Noose Method Veka.a About Noose te Selec Cf: Ev Inc of Cf: Ev Inc of d Start	wek tributeSelection.CfsSub: sSubsetEval : aluates the worth of a su induator performed and redundancy between the debug oNotCheckCapabilities locallyPredictive missingSeparate numThreads	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree em. False False True False 1	More Capabilities	
Prep Attribu	te Evaluator Noose Method Veka.a About Noose te Selec Cf: Ev Inc of Cf: Ev Inc of d Start	wek ttributeSelection.CfsSub sSubsetEval : aluates the worth of a su dividual predictive ability of redundancy between the debug oNotCheckCapabilities locallyPredictive missingSeparate	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree em. False False True False 1	More Capabilities	
Prep Attribu	te Evaluator Noose Method Nethod Veka.a About Ct: te Selec Dross-va Itrend Start list (righ	wek tributeSelection.CfsSub: sSubsetEval : aluates the worth of a su induator performed and redundancy between the debug oNotCheckCapabilities locallyPredictive missingSeparate numThreads	Weka Explorer Select attributes Visualize ka.gui.GenericObjectEditor osetEval ubset of attributes by considering the of each feature along with the degree em. False False True False 1	More Capabilities	

Inside the weka explorer window there are six tabs:

1. **Preprocess-** used to choose the data file to be used by the application.

Open File- allows for the user to select files residing on the local machine or recorded medium **Open URL**- provides a mechanism to locate a file or data source from a different location specified by the user

Open Database- allows the user to retrieve files or data from a database source provided by user

2 Classify- used to test and train different learning schemes on the preprocessed data file under experimentation

		Weka Explorer		- 0 ×	
process Classify Cluster Associate	Select attributes Visualize				
fier					
🗋 weka					
Classifiers	10 2000				
 bayes functions 	sifier output			 	
 Introduction Introduction 					
+ 🚞 meta					
► 🔐 misc ▼ 🔐 rules					
DecisionTable					
S JRip MtRules					
OneR					
PART .					
ZeroR.					
· w sees					
5	Glose				
				Log 💉 x0	
	1 😜 🛃 🖏 🕅	ç 💷 💌 🔬 🕯	🖻 🏧 🗴 📜		

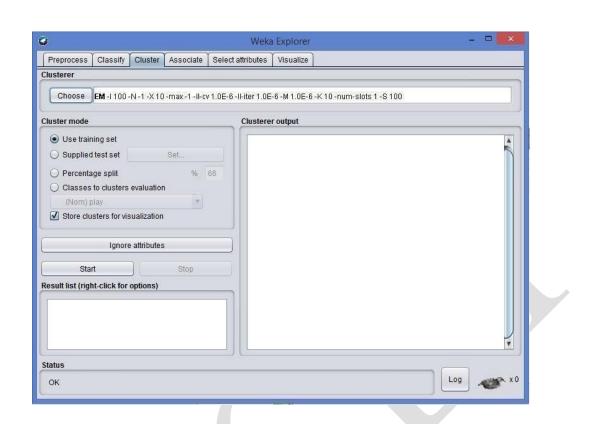
Fig: 1.3 choosing Zero set from classify

Again there are several options to be selected inside of the classify tab. Test option gives the user the choice of using four different test mode scenarios on the data set.

- 1. Use training set
- 2. Supplied training set
- 3. Cross validation
- 4. Split percentage

3. Cluster- used to apply different tools that identify clusters within the data file.

The Cluster tab opens the process that is used to identify commonalties or clusters of occurrences within the data set and produce information for the user to analyze.



4. Association- used to apply different rules to the data file that identify association within the data. The associate tab opens a window to select the options for associations within thedataset.

ø				Weka	Explorer		
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize		
Associator							
Choose	Apriori -N	10-T0-C	0.9 -D 0.05 -	U 1.0 -M 0.1 -S -1.0	-c -1		
	4.00	ociator ou	tout				
Start St	op AS		itput				
Result list (rig)	nt						
-							
Status						_	
ок						L	og po
UN							

5. Select attributes-used to apply different rules to reveal changes based on selected attributes inclusion or exclusion from the experiment

6. Visualize- used to see what the various manipulation produced on the data set in a 2D format, in scatter plot and bar graph output.

2. Experimenter - this option allows users to conduct different experimental variations on data sets and perform statistical manipulation. The Weka Experiment Environment enables the user to create, run, modify, and analyze experiments in a more convenient manner than is possible when processing the schemes individually. For example, the user can create an experiment that runs several schemes against a series of datasets and then analyze the results to determine if one of the schemes is (statistically) better than the other schemes.

Open	Save	New
Results Destination		
ARFF file Tilename:		Browse
Experiment Type	Iteration Control	
Cross-validation	Number of repetitions:	
Number of folds:	 Data sets first 	
Classification Regression	O Algorithms first	
Datasets	Algorithms	
Add new Edit selected Delete selected	ted Add new Edit	selected Delete selected
Up Down	Load options S	ave options Up Down

Results destination: ARFF file, CSV file, JDBC database.

Experiment type: Cross-validation (default), Train/Test Percentage Split (data randomized).

Iteration control: Number of repetitions, Data sets first/Algorithms first.

Algorithms: filters

3. Knowledge Flow -basically the same functionality as Explorer with drag and drop functionality. The advantage of this option is that it supports incremental learning from previous results

4. Simple CLI - provides users without a graphic interface option the ability to execute commands from a terminal window.

b. Explore the default datasets in weka tool.

Click the "*Open file*..." button to open a data set and double click on the "*data*" directory. Weka provides a number of small common machine learning datasets that you can use to practice on. Select the "*iris.arff*" file to load the Iris dataset.

File Home Share Vi	ew				~
-) -> - ↑ 🌗 ト This PC	► OS (C:) ► Program Files ► Weka-3-8 ► data		~ C	Search data	م
😽 Favorites	Name	Date modified	Туре	Size	
Desktop	airline.arff	4/14/2016 8:28 AM	ARFF Data File	3 KB	
Downloads	G breast-cancer.arff	4/14/2016 8:28 AM	ARFF Data File	29 KB	
🖳 Recent places	ontact-lenses.arff	4/14/2016 8:28 AM	ARFF Data File	3 KB	
ConeDrive	🜍 cpu.arff	4/14/2016 8:28 AM	ARFF Data File	6 KB	
	cpu.with.vendor.arff	4/14/2016 8:28 AM	ARFF Data File	7 KB	
🜏 Homegroup	G credit-g.arff	4/14/2016 8:28 AM	ARFF Data File	159 KB	
	diabetes.arff	4/14/2016 8:28 AM	ARFF Data File	37 KB	
🖳 This PC	glass.arff	4/14/2016 8:28 AM	ARFF Data File	18 KB	
he Desktop	Appothyroid.arff	4/14/2016 8:28 AM	ARFF Data File	304 KB	
Documents	ionosphere.arff	4/14/2016 8:28 AM	ARFF Data File	79 KB	
🚺 Downloads	iris.2D.arff	4/14/2016 8:28 AM	ARFF Data File	4 KB	
Music	🥥 iris.arff	4/14/2016 8:28 AM	ARFF Data File	8 KB	
E Pictures	labor.arff	4/14/2016 8:28 AM	ARFF Data File	9 KB	
📓 Videos	ReutersCorn-test.arff	4/14/2016 8:28 AM	ARFF Data File	501 KB	
Los (C:)	ReutersCorn-train.arff	4/14/2016 8:28 AM	ARFF Data File	1,198 KB	
👝 New Volume (F:)	ReutersGrain-test.arff	4/14/2016 8:28 AM	ARFF Data File	501 KB	
Rew Volume (G:)	ReutersGrain-train.arff	4/14/2016 8:28 AM	ARFF Data File	1,198 KB	
	segment-challenge.arff	4/14/2016 8:28 AM	ARFF Data File	196 KB	
📬 Network	Segment-test.arff	4/14/2016 8:28 AM	ARFF Data File	108 KB	
SAIKRISHNAN	Soybean.arff	4/14/2016 8:28 AM	ARFF Data File	199 KB	
	supermarket.arff	4/14/2016 8:28 AM	ARFF Data File	1,979 KB	

Fig: 1.7 Different Data Sets in weka

Exercise:

1. Normalize the data using min-max normalization



Signature of the Faculty

Experiment 2: Creating new ARFF file

Aim: Creating a new ARFF file

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software in WEKA, each data entry is an instance of the java class weka.core. Instance, and each instance consists of a For loading datasets in WEKA, WEKA can load ARFF files. Attribute Relation File Format has two sections:

- 1. The Header section defines relation (dataset) name, attribute name, and type.
- 2. The Data section lists the data instances.

	n2 jap 🛛 블 profile jap 🗶 🚆 homepage html 🖾 📇 login jap 🗶 💾 new 3 jap 🖄 🔚 index 3 html 🔀 블 config.rb 🔀 🗮 index ja 🕉 🚍 rat jap 🖄
_	
1	grelation weather.symbolic
2	
3	<pre>@attribute outlook {sunny, overcast, rainy}</pre>
4	<pre>@attribute temperature {hot, mild, cool} @attribute humidity {high, normal}</pre>
5	<pre>@attribute humidity {high, hormal; @attribute windy {TRUE, FALSE}</pre>
7	<pre>@attribute windy {ikoe, rabse; @attribute play {yes, no}</pre>
8	Catchibace play (yes, no)
9	edata
10	sunny, hot, high, FALSE, no
11	sunny, hot, high, TRUE, no
12	overcast, hot, high, FALSE, yes
13	rainy, mild, high, FALSE, yes
14	rainy, cool, normal, FALSE, yes
15	rainy, cool, normal, TRUE, no
16	overcast, cool, normal, TRUE, yes
17	sunny,mild,high,FALSE,no
18	sunny, cool, normal, FALSE, yes
19	rainy,mild,normal,FALSE,yes
20	sunny,mild,normal,TRUE,yes
21	overcast,mild,high,TRUE,yes
22	overcast, hot, normal, FALSE, yes
23	rainy, mild, high, TRUE, no
24	

The figure above is from the textbook that shows an ARFF file for the weather data. Lines beginning with a % sign are comments. And there are three basic keywords:

- \square "@relation" in Header section, followed with relation name.
- \square "@attribute" in Header section, followed with attributes name and its type (or range).
- \Box "@data" in Data section, followed with the list of data instances.

The external representation of an Instances class Consists of:

- A header: Describes the attribute types
- Data section: Comma separated list of data

Besterbalte former standy (Taller, Faller), Besterbalte play (yes, tot) Besterbalte play (yes, tot) Besterbalte play, fort, fight, Faller, ma Besterbalte play, faller, faller, ma Besterbalte play (faller), faller, ma <t< th=""><th>Bretature sealare against un Bretature sealare against un Matribute outlock (punny, oversast, rainy</th><th></th><th>and Hereits</th><th>B Hondie name et 1</th><th>1</th><th></th><th></th><th></th></t<>	Bretature sealare against un Bretature sealare against un Matribute outlock (punny, oversast, rainy		and Hereits	B Hondie name et 1	1			
Bows Control	Gestvinute how mity prigh, normery	2		Gause An		(A).		
 senty, het, Halfsl, Halfs	and the second sec	Street a	Li cus		· 0321	•		
Starte as type - Hownellined Me (Tar) Cancel	servey, het. high, FALSE, no overcome, het. high, FALSE, no overcome, het. high, FALSE, no overcome, het. high, FALSE, yes rainy, col, normal, FALSE, yes rainy, col, normal, FALSE, yes rainy, col, normal, FALSE, yes rainy, col, normal, FALSE, yes rainy, and, high, FALSE, yes rainy, alld, normal, FALSE, yes rainy, alld, normal, FALSE, yes overcome, if (a two, rain), yes overcome, hot, normal, FALSE, yes rainy, hot, normal, FALSE, yes rainy, hot, normal, FALSE, yes	Deserptices Destrop Librario Ra, FC	O atracti O areas an O contact is O quark O quark O and quark O and quark O another O another O another O another O contact is O another O contact is O another O anot	uusat ediscat I Gatt Satt	4/14/3016/64214 4/14/3016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214 4/14/2016/64214	AC ARF#(AC AC AC AC AC AC AC AC AC AC AC AC AC AC	ARF	
			Sites in Spa	Normalitiest Me ("str)	v	Cencel		

References:

https://www.cs.auckland.ac.nz/courses/compsci367s1c/tutorials/IntroductionToWeka.pdf

Exercise:

1. Creating a sample dataset for supermarket (supermarket.arff)



Signature of the Faculty

Experiment 3: Data Processing Techniques on Data Set

Aim: 3a) Pre-process a given dataset based on Attribute selection

To search through all possible combinations of attributes in the data and find which subset of attributes works best for prediction, make sure that you set up attribute evaluator to "Cfs Subset Val" and a search method to "Best First". The evaluator will determine what method to use to assign a worth to each subset of attributes. The search method will determine what style of search to perform. The options that you can set for selection in the "Attribute Selection Mode" fig no: 3.2

1. Use full training set. The worth of the attribute subset is determined using the full set of training data.

2. Cross-validation. The worth of the attribute subset is determined by a process of cross-validation. The "Fold" and "Seed" fields set the number of folds to use and the random seed used when shuffling the data.

Specify which attribute to treat as the class in the drop-down box below the test options. Once all the test options are set, you can start the attribute selection process by clicking on "Start" button.

3			Weka	Explorer		-	• ×
Preprocess Classify	Cluster	Associate	Select attributes	Visualize			
Classifier							
Choose ZeroR							
Test options		Class	ifier output				
Use training set Supplied test set Cross-validation Free Percentage split More options. (Nom) play Start Result list (right-click for options)	% 66 Stop						•
Status							
OK						Log	XI A

Fig: 3.1 choosing Cross validation

When it is finished, the results of selection are shown on the right part of the window and entry is added to the "Result list".

Plot Matrix sno price speed hd ram screen trend ads PlotSize: [100] PointSize: [11] Jiter: Colour: trend (Num) Class Colour 1 1 18 335	Preprocess CI	assify Cluster A	ssociate Select attrib	outes Visualiz			
trend ads PlotSize: [100] PointSize: [1] Jitter: Colour: trend (Num) Class Colour	Plot Matrix	sno	price	speed	1	ram	110000000000000000000000000000000000000
ads	trend	Conner and the second					Ď
PlotSize: [100]	ads						1: :
PlotSize: [100] Update PointSize: [1] Update Jitter: Select Attributes Colour: trend (Num) SubSample %: 7.99 Class Colour	**						
PointSize: [1]	PlotSize: [100]	0		Fast s	crolling (uses more m	iemory)	
Colour: trend (Num) SubSample %: 7.99	PointSize: [1]			Up	date		
Class Colour	Jitter: 🕥			Select /	Attributes		
	Concertain	um)		SubSa	mple % : 7.99		
	Colour: trend (Nu		1.1		57 (UNS		
1 18 35							

2. Visualizing Results

Fig: 3.2 Data Visualization

WEKA"s visualization allows you to visualize a 2-D plot of the current working relation. Visualization is very useful in practice; it helps to determine difficulty of the learning problem. WEKA can visualize single attributes (1-d) and pairs of attributes (2-d), rotate 3-d visualizations (Xgobi-style). WEKA has "Jitter" option to deal with nominal attributes and to detect "hidden" data points.

٥					Weka	a Explore	ŕ					- 0	×
Preprocess	Classify	Cluster	Associate	Select att	tributes	Visualize							
Plot Matrix		sno	pri	ce	s	peed	M	hd	1	ram		screer	
trend										All Strategy and	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1		Ć
ads													
						-							
PlotSize: [100]		0				Alexandra and a second second		ses more r	nemory)				
PointSize: [1]	0					Upd		J					
Jitter:	c		0		-	Select At	tributes	J					
Colour: trend	(Num)				V	SubSam	iple % :	7.99					
Class Colour													
		_	_	_	_	1	-	_		_	_	_	
1						18							35
Status											_	_	
											Lo	100	🔿 x 0

Fig 3.3: Preprocessing with jitter

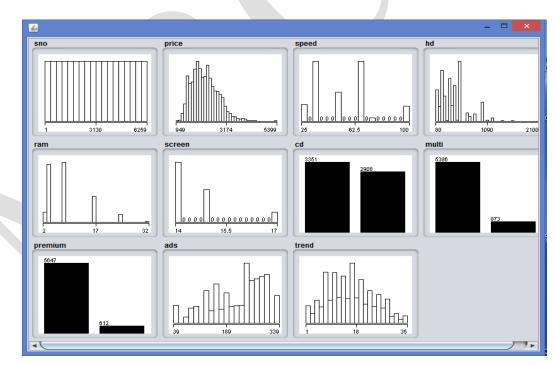


Fig: 3.3 Data visualization

Exercise

1. Explain data preprocessing steps for heart disease dataset.

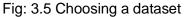
Aim: B. Pre-process a given dataset based on Handling Missing Values

Process: Replacing Missing Attribute Values by the Attribute Mean. This method is used for data sets with numerical attributes. An example of such a data set is presented in fig no: 3.4

-																							
-	<u>5</u>									Con	nputer2	.csv - Exce	1									? 🗹 –	
FILE		E INS	SERT	PAGE LAYOU	T FOR	VULAS	DATA	REVIEW	VIEW													SAI KRIS	HNAN
	K Cut Copy	•	Calibri	Ŧ	11 - A				rap Text					<i>≠</i>			€ ■■	*		∑ AutoSum ↓ Fill -	ΖY		
	✓ Format		BI	<u>U</u> - 🖽 -	- 🖄 - 📕	<u>-</u> = :	= = (e	₩ 🗄 M	erge & Cente	er 👻 🕴	5 - %	• • • • • •	.00 •.0	Conditional Fo	ormatas (Table - St	Cell des v	Insert •	Delete F	ormat	🛃 Clear 🕶	Sort & F Filter ▼ S		
	Clipboard	G.		Font		5		Alignment		5		ımber	5	2	/les			Cells		Ed	iting	cicci	
B12	-	= >	< 🗸	f_x ?																			
	A	в	с	D	E	F	G	н	1	J		к	L	м	N	C		Р	Q	R	S	т	
1 sn	o pr	ice	speed	hd	ram	screen	cd	multi	premium	ads	tre	end											
2	1	1499	2	25 80) 4	4 1	4 no	no	yes		94	1											
3	2	1795	3	3 85	5	2 1	.4 no	no	yes		94	1											
4	3 ?		2	25 170) 4	4 1	.5 no	no	yes		94	1											
5	4	1849	2	25 170) (B 1	4 no	no	no		94	1											
5	5	3295	3	33 340) 1	5 1	.4 no	no	yes		94	1											
7	6	3695	6	6 340		5 1	.4 no	no	yes		94	1											
3	7 ?			25 170			4 yes	no	yes		94	1											
9	8	1995		50 85			.4 no	no	yes		94	1											
0	9	2225		50 210			.4 no	no	yes		94	1											
1	10	2575	_	50 210			.5 no	no	yes		94	1											
12	11 ?			33 170			.5 no	no	yes		94	1											
13	12	2605		6 210			.4 no	no	yes		94	1											
4	13	2045		50 130			.4 no	no	yes		94	1											
5	14	2295		25 245			4 no	no	yes		94	1											
6	15	2699		0 212			4 no	no	yes		94	1											
7 8	16 17	2225 1595		50 130			4 no 4 no	no	yes		94	1											
9	17	2325		3 85 3 210			4 no 5 no	no	yes		94 94	1											
0	18	2325		33 210			.5 no	no	yes yes		94 94	1											
1	20	4395		5 250 56 452			4 no	no	ves		94	1											
22	20	4595		33 130			4 no	no	yes		94	1											
23	22	2795		5 130 56 130			4 no	no	ves		94	1											
	22	2755		10 130			4		yes		04												

Fig: 3.4 Missing values





In this method, every missing attribute value for a numerical attribute is replaced by the arithmetic mean of known attribute values. In Fig, the mean of known attribute values for Temperature is 99.2, hence all missing attribute values for Temperature should be replaced by The table with missing attribute values replaced by the mean is presented in fig. For symbolic attributes Headache and Nausea, missing attribute values were replaced using the most common value of the Replace Missing Values.

Open file Open URL Open DB Ge	enerate Undo	Edit Save	
	Undo	Edit	
filter			
Choose ReplaceMissingValues		Apply	
Current relation	Selected attribute		
Relation: Computer2 Attributes: 11	Name: price	Type: Numeric	
Instances: 6259 Sum of weights: 6259		istinct: 808 Unique: 202 (3%)	
Attributes	Statistic Minimum	Value 949	
All None Invert Pattern	Maximum Mean	5399 2219.76	
	StdDev	580.855	
No. Name		B	2
1 sno 2 ✓ price	Class: trend (Num)	Visualize Al	Ŀ
3 speed 4 hd			
5 ann 6 screen			
7 cd		75	
Remove			
		<u>456311102003</u> 3174 5399	1 2
Status	***		
OK	- Eveloror		
Weka Preprocess Classify Cluster Associate Select attributes	a Explorer	- •	
Weka Preprocess Classify Cluster Associate Select attributes Open tile Open URL Open DB Ge			
Veka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter	Visualize	- • • • • • • • • • • • • • • • • • • •	
Weka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues	Visualize	- •	
Veka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge itter Choose ReplaceMissingValues current relation	Visualize enerate Undo Selected attribute	Edit Save	
Veka Preprocess Classify Cluster Associate Select attributes Open file, Open URL Open DB Ge itter Choose ReplaceMissingValues	Visualize enerate Undo Selected attribute Name: price	- • • • • • • • • • • • • • • • • • • •	
Weka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues Choose ReplaceMissingValues current relation Attributes: 11 Instances: 6259 Sum of weights: 6259	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic	Edit Save Apply istinct 809 Unique: 202 (3%) Value	
Veka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues current relation Relation: Computer2-weka.filters Attributes: 11 Instances: 6259 Sum of weights: 6259 utributes	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D	Edit Save Apply Type: Numeric istinct: 809 Unique: 202 (3%)	
Weka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues Choose ReplaceMissingValues current relation Attributes: 11 Instances: 6259 Sum of weights: 6259	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Mean	Edit Save Edit Save Apply istinct 809 Unique: 202 (3%) Value 949 5399 2219.76	
Weka Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Get itter Choose ReplaceMissingValues Choose ReplaceMissingValues current relation Attributes: 11 Instances: 6259 ttributes All None No. Name	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Maximum	Edit Save Edit Save Apply istinct 809 Unique: 202 (3%) Value 949 5399	
Wekz Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues Choose ReplaceMissingValues current relation Attributes: 11 Instances: 6259 Sum of weights: 6259 ttributes All None All None Invert No. Name 1 1 sno 1	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Mean	Edit Save Edit Save Apply istinct 809 Unique: 202 (3%) Value 949 5399 2219.76	
Wekz Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge itter Choose ReplaceMissingValues current relation Relation: Computer2-weka filters Attributes: 11 Instances: 6259 Sum of weights: 6259 xttributes All None Invert Pattern No. Name 1 sno 4 1 sno 4 hd 4	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Maximum Mean StdDev	Edit Save Edit Save Apply Type: Numeric istinct: 809 Unique: 202 (3%) Value 949 5399 2219.76 580.716	
Wekk Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge ilter Choose ReplaceMissingValues Current relation Relation: Computer2-weka.filters Attributes: 11 Instances: 6259 Sum of weights: 6259 Attributes Attributes Attributes: 11 Instances: 6259 Sum of weights: 6259 Attributes Attributes Attributes Attributes Force Force Attributes Force Force Force Attributes Force Force Force Force	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Maximum Mean StdDev	Edit Save Edit Save Apply Type: Numeric istinct: 809 Unique: 202 (3%) Value 949 5399 2219.76 580.716	
Wekz Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge iller Choose ReplaceMissingValues Current relation Attributes: 11 Instances: 6259 Sum of weights: 6259 Xttributes Attributes: 11 No. Name 1 sno 1 sno 4 hd	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Maximum Mean StdDev	Edit Save Edit Save Apply Type: Numeric istinct: 809 Unique: 202 (3%) Value 949 5399 2219.76 580.716	
Wekz Preprocess Classify Cluster Associate Select attributes Open file Open URL Open DB Ge Choose ReplaceMissingValues Current relation Relation: Attributes: 11 Instances: 6259 Sum of weights: 6259 All None Invert Pattern No. Name 4 All Attributes 1 sno 4 Attributes 4 6 screen 6 screen 4	Visualize enerate Undo Selected attribute Name: price Missing: 0 (0%) D Statistic Minimum Maximum Mean StdDev	Edit Save Edit Save Apply Type: Numeric istinct: 809 Unique: 202 (3%) Value 949 5399 2219.76 580.716	

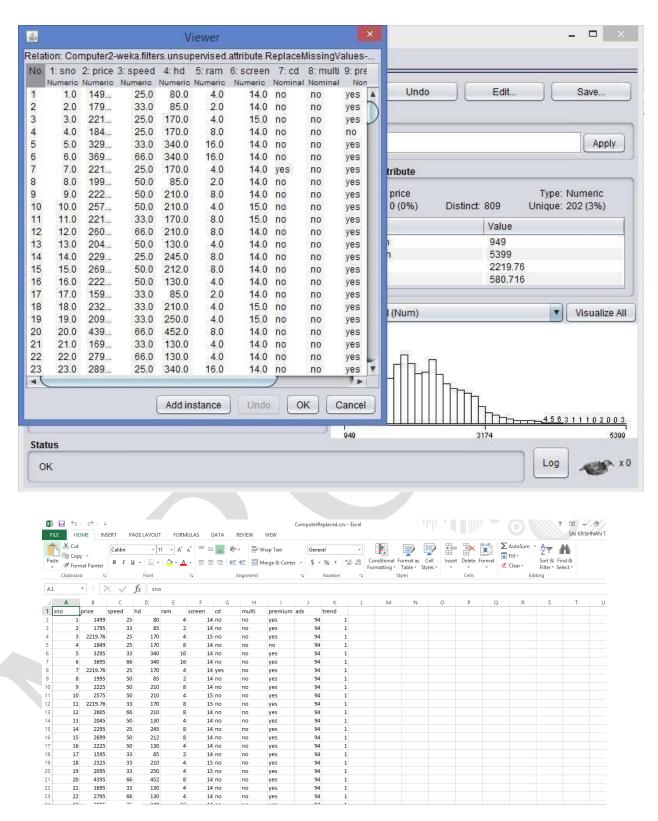


Fig: 3.6 Replaced values

Exercise

1. Create your own dataset having missing values included and perform preprocessing



Signature of the Faculty

Experiment 4: Data cube construction – OLAP operations

An OLAP cube is a term that typically refers to multi-dimensional array of data. OLAP is an acronym for online analytical processing,[1]which is a computer-based technique of analyzing data to look for insights. The term cube here refers to a multi-dimensional dataset, which is also sometimes called a hypercube if the number of dimensions is greater than 3.

Operations:

1. **Slice** is the act of picking a rectangular subset of a cube by choosing a single value for one of its dimensions, creating a new cube with one fewer dimension.[4] The picture shows a slicing operation: The sales figures of all sales regions and all product categories of the company in the year 2005 and 2006 are "sliced" out of the data cube.

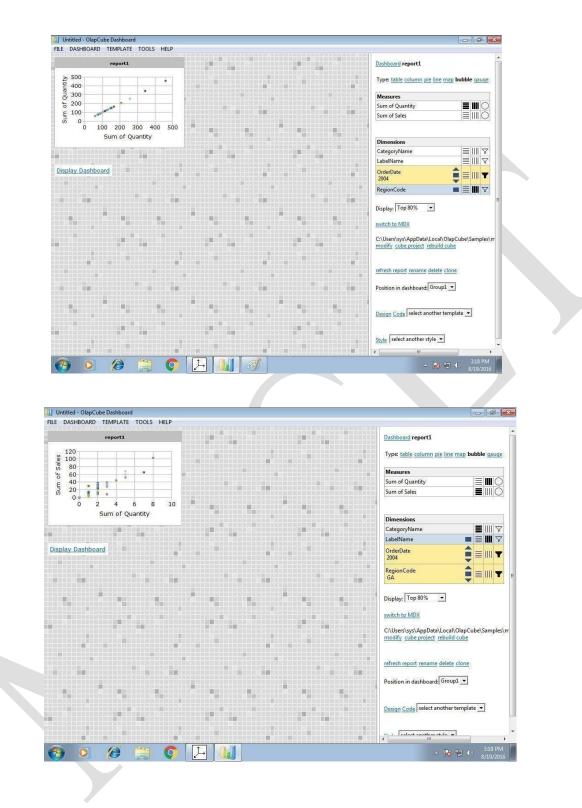
2. **Dice:** The dice operation produces a subcube by allowing the analyst to pick specific values of multiple dimensions.[5]The picture shows a dicing operation: The new cube shows the sales figures of a limited number of product categories, the time and region dimensions cover the same range as before.

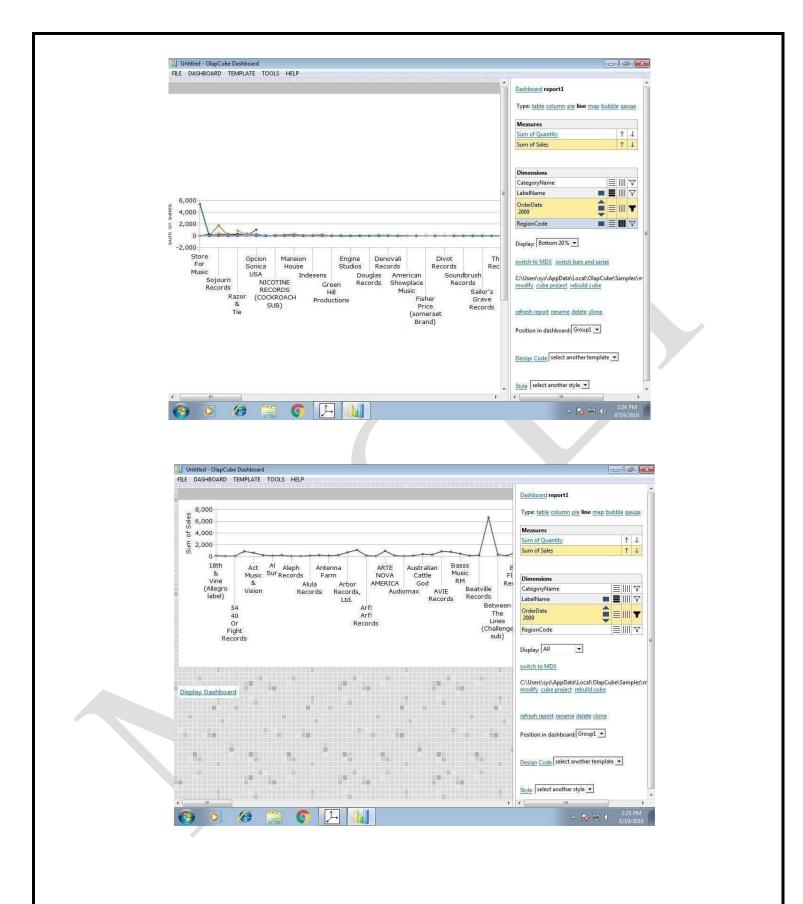
3. **Drill Down/Up** allows the user to navigate among levels of data ranging from the most summarized (up) to the most detailed (down).[4] The picture shows a drill-down operation: The analyst moves from the summary category "Outdoor-Schutzausrüstung" to see the sales figures for the individual products.

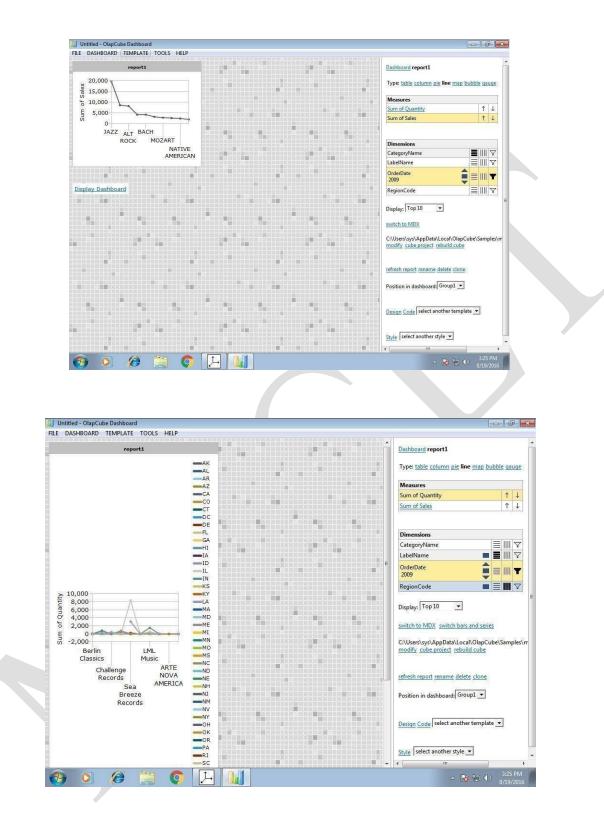
4. **Roll-up:** A roll-up involves summarizing the data along a dimension. The summarization rule might be computing totals along a hierarchy or applying a set of formulas such as "profit = sales - expenses".

5. **Pivot** allows an analyst to rotate the cube in space to see its various faces. For example, cities could be arranged vertically and products horizontally while viewing data for a particular quarter. Pivoting could replace products with time periods to see data across time for a single product.

FILE	HOME	CREATE	EXTERNAL DAT	A DATA	ABASE TOOLS	FIELDS	TABLE						SAI KR
M	- Xc		1.1		Y Selection *				Replace	Accil Calibri	- 11 - E E	· · · · · · · · · · · · · · · · · · ·	-
View	Paste				Advanced -	Refresh		Find		Size to Switch B I U	A - 2 - 8 - = = =	- 12A - 100	B
Views	Clipbo	ormat Painter	2,∉ Rei	move Sort Sort & Filte	Toggle Filter	All - XI	Delete - E	More * 🗟 : Find	Select * F	it Form Windows * Window	Text Formatting		
		bje… ⊛ «		Soft & Thite	3		Records	Fillo	2	Window	Text Formatting		(9
Search.	ICCESS OI	DJE											
Tables		*					1		Albums				- 1
	lbums ustomers		Alt	oumID • 5202	Album Josh Groban	Name	• Groban,	Artist Josh	*	LabelName Reprise	CategoryName POP		• Rele
	mployees		۲	5306	Shaman's Hea		Shastro			Malimba Records	RELAXATION	1	1.26
	orderDetails				Leaves Of Hop Russian Soul	pe	Ginkgo G	iarden Daniil: vc		BSC Music The Classics Label (Launch sub)	NEW AGE CELLO		13.2 5.19
0	Orders		۲	5310	Live At The Sa		Chimps,	The		MuSick Recordings, Inc.	ALT ROCK	1:	1.19
					Toddler Favor Celebration o			r Little People elebration Series		Music For Little People Music For Little People	CHILDREN		4.72 0.12
					Futureperfect		VNV Nat			Metropolis Records	ALT ROCK		13.2
			•		Joined at the Schumann's B		Williams Yasuda, I			TCB Music Winter & Winter	JAZZ JAZZ		13.2 5.99
			œ.		i Splay			n/Alasnoaxis		Winter & Winter	JAZZ		5.99
					Indian Journe		Voyager			Columbia River Entertainment	INDIA		5.59
					Bedtime Lulla Journey to the		Various /			Big Blue Dog DOMO Records	CHILDREN NEW AGE		5.99 5.19
				5333	Soul of the Be	eloved	Silbey, U	lma		DOMO Records	NEW AGE	1	5.19
					Piano for Love Piano by Requ		Various /			Columbia River Entertainment Columbia River Entertainment	EASY LISTENING EASY LISTENING		1.19 1.19
				5338	Encore une va	alse ?	Nord Sug	e e e e e e e e e e e e e e e e e e e		Buda Musique	AFRICA		13.2
					Piano Giants a Boli Bana	at Bob Haggar	rt' Haggart, Djembe,	Hyman, Smith, Sutto Tahirou		Arbors Records Playasound	JAZZ AFRICA		8.8 13.2
			۲	5347	"Wanderer" F		ata Kruchin,	Semion: piano		Meridian	SCHUBERT		1.39
			Record: H	5348	isea of Tranqu	uility: Relaxati X No Filter Se		S WELLNESS SERIES		Lifestyles	NEW AGE		3.5
Datashee	et View		-										NUM D
	E			1		🌄 🟹		W1 🕹		🞽 X 🗄 🔑			al 🔯 🚺
CURF Cube: m Dime	nusic ensions CategoryNa Categor	RY QUERIA ame ryName		Albur Albur	open cube 1 ms mID	Order 7 Order	rDetails rID	<u>rs new relationshi</u>	<u>o</u> refresh	a database properties new Customers CustomerID CustomerName	query	[
CURF Cube: m Dime 2 12, C	RENT QUEF nusic ensions CategoryNa Categor abelName LabelNa Artist	RY QUERIA ame ryName ame		Albur Albur Albur Albur Artist Label	open cube f ms mID mName t IName	Order	rDetails rID mID Price	Orders	2 refresh	Customers CustomerID	query	[
CURF Cube: m Dime Cube: m Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor abelName LabelNa	RY QUERIA ame ryName ame	bu	Albur Albur Albur Albur Artist Label	open cube (ms mID mName t IName goryName Price	Order Order Albur UnitP	rDetails rID mID Price]	2 refresh	Customers CustomerID CustomerName RegionCode	query		
CURF Cube: m Dime 2 12, C 2 12, L 3 12, L 4 10 10 10 10 10 10 10 10 10 10 10 10 10 1	RENT QUEF nusic ensions CategoryNa CategoryNa Categor abelName LabelNa Artist AlbumN OrderDate Year Month r	RY QUERIE ame ryName ame Name name	<u>bu</u>	Albur Albur Albur Artist Label Categ UnitP	open cube { ms mID mName t IName goryName rrice	Order Order Albur UnitP Quan	rDetails rID mID trice ttity	Orders OrderID CustomerID		Customers CustomerID CustomerName RegionCode Country Employees			
	RENT QUEF nusic ensions CategoryNa CategoryNa CategoryNa CategoryNa LabelName LabelName LabelName Artist AlbumN OrderDate Year Month r RegionCode	RY QUERIE ame ryName ame Name name	<u>bu</u> ≮ Pre	Albur Albur Albur Albur Albur Artist Label Categ UnitP	open cube { ms miD mName t Name price mi Albums	Order Order Albur UnitP Quan	rDetails rID mID trice ttity	Orders OrderID CustomerID pen cube refresh		Customers CustomerID CustomerName RegionCode Country Employees	nove table rename tabl	e change t	able
CURF Cube: m Dime L C L L L L L L L L L L L L L R L L Meas	RENT QUEF nusic ensions CategoryNa CategoryNa CategoryNa CategoryNa LabelName LabelName LabelName Artist AlbumN OrderDate Year Month r RegionCode	RY QUERIF ame ry/Name sime Name name e	< Alternational	Albur Albur Albur Albur Artist Label Categ UnitP	open cube { ms miD mName t Name price mi Albums	Order Order Albur UnitP Quan	rDetails rID mID rice titity d cube	Orders OrderID CustomerID pen cube refresh		Customers CustomerID CustomerName RegionCode Country Employees Employees	nove table rename tabl	e change t	<u>able</u>
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	<u>bu</u> ≮ Pre	Albur Albur Albur Albur Albur Artist Label Categ UnitP eview of bumID Al 03 Jo	open cube (ms mID mName c Name goryName Price "" Albums <u>lbumName</u> osh Groban	Order Order Albur UnitP Quan	rDetails rID mID rice titity d cube	Orders OrderID CustomerID pen cube refresh st ban, Josh		Customers CustomerID CustomerName RegionCode Country Employees	nove table rename tabl	e change t	able e Release 4/23/200
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa CategoryNa CategoryNa LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCodi sures Sum of Qua	RY QUERIN ame ryName ame Name e name e antity es	د ۲ ۲ ۲ ۲ ۲ ۲ ۲ ۲ ۲	Albur Albur Albur Artist Label Categ UnitP eview of burnID Al 33 Jo 26 St	open cube { ms miD mName t Name price mi Albums	Order Order Albur UnitP Quan	rDetails rID mID trice titty d cube Groi Sha:	Orders OrderID CustomerID pen cube refresh st ban, Josh		Customers CustomerID CustomerName RegionCode Country Employees a show/hide columns rer LabelName Reprise	nove table rename tabl CategoryName POP	e change t UnitPric 5.39	able e Release[4/23/200 4/25/200
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alt 530 530 530 	Albur Albur Albur Artist Label Categ UnitP eview of 5 burnID Ar 03 Jo 06 Sh 07 Le	open cube miD miD mName ! Name goryName /rice !! Albums ! bumName osh Groban haman's Hea aeaves Of Hop	Order Order Albur UnitP Quan	rDetails rID mID trice titity d cube Arti Grol Sha: Ginl	Orders OrderID CustomerID pen cube refresh stro stro cgo Garden		Customers CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun	nove table rename tabl CategoryName POP RELAXATION NEW AGE	e change t UnitPric 5.39 11.26 13.2	able c Released 4/23/200 4/27/200 4/27/200
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 4 4 4 4 4 530 530 530 530 530 530 531 	Albur Albur Albur Artist Label Categ UnitP Eview of DaumID Al D3 Jo D6 SH D7 Le D9 Rt	open cube miD miD mName illuame goryName irice illuams ilbumName osh Groban haman's Hea eaves Of Hop ussian Soul	Order Order Albur UnitP Quan	rDetails rID mID trice titity d cube Grod Sha: Ginl Sha	Orders OrderID CustomerID CustomerID st ban, Josh stro kgo Garden fran, Daniil: vc		Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees a show/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub)	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO	e change t UnitPric 5.39 11.26 13.2 15.19	able ReleaseE 4/23/200 4/27/200 5/1/2002
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 4 4 4 4 530 530 531 	Albur Albur Albur Albur Artist Label Categ UnitP Eview of DoumID Al D3 Jo D6 SH D7 Le D9 Ri L0 Li	open cube miD miD mName i liName goryName rrice III Albums libumName ash Groban haman's Hea eaves Of Hop ussian Soul vive At The Sa	Order Order Albur UnitP Quan	rDetails rID mID trice titity d cube Grod Sha: Ginl Sha! Chin	Orders OrderID CustomerID ban, Josh stro cgo Garden fran, Daniil: vc mps, The	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc.	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK	e change t UnitPrice 5.39 11.26 13.2 15.19 11.19	able e Release 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alt 530 530 530 531 533 533 533 533 533 	Albur Albur Albur Artist Label Categ UnitP eview of bumID Al 03 Jo 06 St 07 Le 09 Ru 10 Li 11 Te	open cube miD miD miD miName tiliname trice ilinamaris Hea caves Of Hop ussian Soul ve At The Sa oddler Favori	Order Order Albur UnitP Quan	rDetails rID mID brice titity d cube Grol Sha Ginl Sha Chin Mus	Orders OrderID CustomerID Den cube refresh stro stro stro stro fran, Daniil: vc mps, The sic For Little People	hide dat	Customers CustomerlD CustomerlName RegionCode Country Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc.	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72	able e ReleaseII 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/1/2002
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alti 530 533 	Albur Albur Albur Artist Label Categ UnitP eview of burnID Al 03 Jo 06 Sh 07 Le 09 Rt 10 Li 11 To 12 Co	open cube mID mID mIVame goryName trice mi Albums <u>IbumName</u> osh Groban haman's Hea eaves Of Hop ussian Soul vice At The Sa oddler Favori elebration of	Order Albur UnitP Quan	rDetails rID mID rrice tity d cube co sha sha Sha Sha Chin Mus Chin	Orders OrderID CustomerID ban, Josh stro cgo Garden fran, Daniil: vc mps, The	hide dat	Customers CustomerID CustomerName RegionCode Country Employees Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc. MuSick Recordings, Inc.	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN CHILDREN	e change 1 UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12	able Released 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alt 530 530 530 531 533 533 533 533 533 	Albur Albur Albur Albur Artist Label Categ UnitP Eview of burnD Al 03 Jo 06 St 07 Le 09 R 10 Li 11 Tr 12 Cr 22 Ft	open cube miD miD miD miName tiliname trice ilinamaris Hea caves Of Hop ussian Soul ve At The Sa oddler Favori	Order Albur UnitP Quan	rDetails rID mID vrice titty d cube Grol Sha Ginl Sha Ginl Chin Mus Chin	Dirders OrderID CustomerID CustomerID Stro Go Garden fran, Daniil: vc mps, The ci: For Little People d's Celebration Se	hide dat	Customers CustomerlD CustomerlName RegionCode Country Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc.	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72	able e Release 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/5/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 <td>Alburn Alburn Alburn Artist Label Categ UnitP UnitP Data Alburn Artist Label Categ UnitP D D D Alburn Alburn Artist Label Categ UnitP D D Alburn Alburn Alburn Artist Label Categ D D D D Alburn Albur</td><td>open cube miD miD miDame : Name : Name : Mibums <u>ibumName</u> osh Groban haman's Hea eaves Of Hop ussian Soul ussian Soul ussian Soul elebaration of utureperfect</td><td>Order Order Albur UnitP Quan Image aling pe afari Club ites Too! f America Hip</td><td>Details HD mID trice tity d cube Chi Chi Chi VNV Will</td><td>Orders OrderID CustomerID CustomerID totan, Josh stro kro Garden fran, Daniil: vc mps, The cic For Little People cid's Celebration Se (Nation</td><td>hide dat</td><td>Customers CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc. Music For Little People Music For Little People Music For Little People Music For Little People</td><td>CategoryName POP RELAXATION NEW AGE chi CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK</td><td>e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2</td><td>able ce Released 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/5/2002 5/7/2002 5/7/2002</td>	Alburn Alburn Alburn Artist Label Categ UnitP UnitP Data Alburn Artist Label Categ UnitP D D D Alburn Alburn Artist Label Categ UnitP D D Alburn Alburn Alburn Artist Label Categ D D D D Alburn Albur	open cube miD miD miDame : Name : Name : Mibums <u>ibumName</u> osh Groban haman's Hea eaves Of Hop ussian Soul ussian Soul ussian Soul elebaration of utureperfect	Order Order Albur UnitP Quan Image aling pe afari Club ites Too! f America Hip	Details HD mID trice tity d cube Chi Chi Chi VNV Will	Orders OrderID CustomerID CustomerID totan, Josh stro kro Garden fran, Daniil: vc mps, The cic For Little People cid's Celebration Se (Nation	hide dat	Customers CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc. Music For Little People Music For Little People Music For Little People Music For Little People	CategoryName POP RELAXATION NEW AGE chi CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2	able ce Released 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/5/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pret Alth 530 531 532 533 534 535 535 535 536 537 537	Alburn Categ UnitP BurnD Alburn Albu	open cube ms mID mName goryName /rice /// Albums ibumName osh Groban haman's Hea eaves Of Hop usian Soul visa Soul sita Association of utareperfect	Order Order Albur UnitP Quan Image aling pe afari Club ites Too! f America Hip	Details ID mID mID mID mID mID mID mID	Orders OrderID CustomerID CustomerID ban, Josh stro kgo Garden fran, Daniik vc mps, The sic For Little People d's Celebration See Nation iams, Buster	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Employees Easthow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc. Music For Little People Music For Little People Metropolis Records TCB Music	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN ALT ROCK JAZZ	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2	able ce ReleaseII 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/5/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alt 530 531 532 533 534 535 535 536 537 537	Album Albuk Albur Date	open cube ms mID mName iName goryName yrice milliame goryName yrice milliame ash Groban haman's Hea eaves Of Hop usian Soul usian Soul sive At The Sa oddler Favori elebration of utureperfect pined at the H chumann's B	Order Order Albur UnitP Quan	Details AD mID trice tity d cube Chi Chi Chi VNV Will Yase Blac	Orders OrderID CustomerID Et ban, Josh stro kgo Garden fran, Daniił: vc mps, The ici: For Little People d's Celebration Se Valtion Suster Jada Fumio	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) Music For Little People Music For Little People Music For Little People Music For Little People Music For Little People Metropolis Records TCB Music Winter & Winter Winter & Winter	categoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK JAZZ JAZZ	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 15.99	able c ReleaseD 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pret Alth 530 530 531 532 532 533 534 534 535 534 535 535	Album Album Albur Data Joan	open cube ms mID mName c Name goryName price mName c Albums IbumName osh Groban haman's Hea eaves Of Hop ussian Soul vise At The Sa oddler Favori elebration of utureperfect bined at the F chumann's B play dian Journey	order Order Albur UnitP Quan Image aling pe afari Club ites Too! f America Hip Bar Music	rDetails rD mID trice tity d cube frice Artii Groi Shai Chii Chii Chii VNV Will Yasi Blac Voy	Orders OrderID CustomerID DustomerID ban, Josh stro kgo Garden fran, Daniil: vc mps, The sic For Little People d's Celebration See / Nation iams, Buster Juda, Fumio k, Jim/Alasnoaxis ager Series	hide dat	Customers CustomerlD CustomerlD CustomerName RegionCode Country Employees Em	nove table rename tabl CategoryName POP RELAXATION NEW AGE cello ALT ROCK CHILDREN ALT ROCK CHILDREN ALT ROCK JAZZ JAZZ JAZZ INDJA	e change t UnitPric 5.39 11.26 13.2 15.19 4.72 10.12 13.2 13.2 13.2 13.2 13.9 13.9 13.9 13.9 15.99	able c Release 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pre Alth 533 534 535 536 537 537	Alburn Catego Outper Main Alburn Albu	open cube ms mID mName goryName /rice /// Albums ibumName osh Groban haman's Hea eaves Of Hop usian Soul visa Soul elebration of utureperfect pined at the F chumann's B	Order Order Albur UnitP Quan Image aling pe afari Club f America Hip Bar Music ys bies	Details rD mID rrice tity Arti Grol Sha Grol Sha Chil Sha Chil Mus Chil WWW Will Yasu Blac Voy Vari	Orders OrderID CustomerID CustomerID State ban, Josh stro kgo Garden fran, Daniil: vc mps, The cic For Little People d's Celebration See Nation iams, Buster Jda, Fumio Jda, Fumio	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) Music For Little People Music For Little People Music For Little People Music For Little People Music For Little People Metropolis Records TCB Music Winter & Winter Winter & Winter	nove table rename tabl CategoryName POP RELAXATION NEW AGE ch CELLO ALT ROCK CHILDREN ALT ROCK JAZZ JAZZ JAZZ	e change 1 UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 13.2 13.9 15.99 15.99	able ReleaseD 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 4 4 4 4 4 53 54 54	Albur Vieworf Go So So So Albur Vieworf Vieworf Albur Vieworf Albur Vieworf Albur Vieworf Albur Vieworf Vieworf Vieworf Vieworf <t< td=""><td>open cube ms mID mName iName goryName yrice iII Albums ibumName osh Groban haman's Hea eaves Of Hop usian Soul visan Soul sisan So</td><td>Albur UnitP Quan unitP Quan saling pe afari Club rites Too! f America Hip Bar Music ys bies e Heart</td><td>China China China</td><td>Orders OrderID CustomerID EustomerID est ban, Josh stro kgo Garden fran, Daniik vc mps, The icic For Little People d's Celebration Se Celebration Se iarns, Buster uda, Fumio ka, Jim/Alasnoaxis ager Series ous Artists</td><td>hide dat</td><td>Customers CustomerlD CustomerlD CustomerlName RegionCode Country Employees Employees Employees Automation for the count Reprise Malimba Records BSC Music Music For Little People Music For Little People Metropolis Records TCB Music Winter & Winter Columbia River Entertainment Big Blue Dog</td><td>Increase Increase Increase</td><td>e change f UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 15.99 15.99 15.99 5.59 6.99</td><td>able ReleaseD 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/</td></t<>	open cube ms mID mName iName goryName yrice iII Albums ibumName osh Groban haman's Hea eaves Of Hop usian Soul visan Soul sisan So	Albur UnitP Quan unitP Quan saling pe afari Club rites Too! f America Hip Bar Music ys bies e Heart	China	Orders OrderID CustomerID EustomerID est ban, Josh stro kgo Garden fran, Daniik vc mps, The icic For Little People d's Celebration Se Celebration Se iarns, Buster uda, Fumio ka, Jim/Alasnoaxis ager Series ous Artists	hide dat	Customers CustomerlD CustomerlD CustomerlName RegionCode Country Employees Employees Employees Automation for the count Reprise Malimba Records BSC Music Music For Little People Music For Little People Metropolis Records TCB Music Winter & Winter Columbia River Entertainment Big Blue Dog	Increase Increase	e change f UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 15.99 15.99 15.99 5.59 6.99	able ReleaseD 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 4 Pret Alt 530 533 534 535 535	Albur Catego DomD Albur Diand Job Son Albur Catego Albur	open cube miD miD miD miDame t lName goryName diversion approximation approximation minimitant approximation app	Order Order Albur UnitP Quan Ining pe fari Club fari Club fari Club fari Music sar Music ys bies e Heart loved	rDetails rD mID rrice tity Arti Grol Sha Grol Sha Grol Sha Chil Mus Chil Mus Chil VNV Will Yast Blac Voy Vari Vari Sha	Orders OrderID CustomerID CustomerID tustome	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) MuSick Recordings, Inc. Music For Little People Metropolis Records TCB Music Winter & Winter Columbia River Entertainment Big Blue Dog DOMO Records DOMO Records DOMO Records	Inovertable rename table CategoryName POP RELAXATION NEW AGE Ch CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK JAZZ JAZZ JAZZ JAZZ JAZZ INDJA CHILDREN CHILDREN	e change t JunitPric 5.39 11.26 13.2 15.19 4.72 10.12 13.2 15.99 15.99 5.59 6.99 15.19	able ReleaseD 4/23/200 4/25/200 4/27/200 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 Pre Alt 533 534 535 535	Album Album Album Album Album Album Album Album Antist Label Catego UnitP exiew of [Diamond Jain	open cube ms mID mName t Name goryName goryName mName t MIbums IbumName osh Groban haman's Hea eaves Of Hop ussian Soul ve At The Sa oddler Favori utureparfect sined at the F chuman's B play adian Journey edtime Lullal bourney to the oul of the Bel iano for Love	order Order Albur UnitP Quan aling pe afari Club ites Too! f America Hip Sar Music ys bies e Heart sloved ers	rDetails rD mID trice titity d cube frice and shat Grol Shat Grol Shat Grol Shat Chin Vov Will Yass Blac Voy Vari Silb Sib	Orders OrderID CustomerID CustomerID to at the stro customerID to at the stro customerID to at the stro customerID to at the stro customerID the s	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Employees Easthow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) Music For Little People Metropolis Records TCB Music Winter & Winter Winter & Winter Winter & Winter Columbia River Entertainment Big Blue Dog DOMO Records DOMO Records DOMO Records	nove table rename tabl CategoryName POP RELAXATION NEW AGE CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK JAZZ JAZZ JAZZ INDIA CHILDREN NEW AGE NEW AGE EASY LISTENING	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 15.99 15.99 5.59 6.99 15.19 15.19 15.19	able Release 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 Prec Alt 530 530 531 533 534 535 535 535 	Album Album Album Album Album Album Album Artist Label Catego UnitP veview of [J3 J3 J33 J6 J7 Le Catego J00 J1 T L2 Fit J2 J02 J2 J33 J4 J22 J23 J24 S26 S13 J23 J33 S2 J33 S37	open cube ms mID mName t Name goryName mName t Name manager mName mName t Name t Name t Manager t Albums <u>IbumName</u> sh Groban haman's Hea eaves Of Hop ussian Soul ussian Soul ussian Soul ussian Soul sussian Soul	order Albur UnitP Quan Jaling pe afari Club farerica Hip Bar Music ys bies e Heart Jolved ers uest	rDetails rD mID trice titity d cube frice and shat Groi Shat Gini Shat Chin Muu Chii VAri Vari Silb Vari Vari	Orders OrderID CustomerID CustomerID total customerID customerID total customerID total customerID total customerID d's Celebration Se (Nation iams, Buster Juda, Fumio ik, Jim/Alasnoaxis ager Series ous Artists ous Artists ous Artists ous Artists	hide dat	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Eashow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) Music For Little People Metropolis Records TCB Music Winter & Winter Winter & Winter Winter & Winter Columbia River Entertainment Big Blue Dog DOMO Records Columbia River Entertainment Columbia River Entertainment	nove table rename tabl CategoryName POP RELAXATION NEW AGE CELLO ALT ROCK CHILOREN ALT ROCK CHILOREN ALT ROCK JAZZ JAZZ JAZZ JAZZ JAZZ JAZZ INDIA CHILDREN NEW AGE NEW AGE EASY LISTENING EASY LISTENING	e change t 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 15.19 11.19 5.59 6.99 15.19 15.	able ReleaseE 4/23/200 4/25/200 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/1/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002 5/7/2002
CURF Cube: m Dime Cube: m Cube: m Cube: m Cube: m Cube: Cube: Cube: Cube: Cube	RENT QUEF nusic ensions CategoryNa Categor LabelName LabelNa Artist AlbumN OrderDate Year Month r RegionCode sures Sure Sure Sale	RY QUERIN ame ryName ame Name e name e antity es	 Pre Alt 533 534 535 535	Album Bab Data Ja	open cube ms mID mName t Name goryName goryName mName t MIbums IbumName osh Groban haman's Hea eaves Of Hop ussian Soul ve At The Sa oddler Favori utureparfect sined at the F chuman's B play adian Journey edtime Lullal bourney to the oul of the Bel iano for Love	Order Order Albur UnitP Quan Ing pe aling pe dfari Club famerica Hip laar Music ys bies et Heart sloved ers uest uest	rDetails rDetails rD rD rice tity rice Arti Grol Sha Ginl Sha Chii Sha Chii VNV Will Vas Blac Voy Vari Silb Vari Silb Vari Silb Vari Silb Vari	Orders OrderID CustomerID CustomerID to at the stro customerID to at the stro customerID to at the stro customerID to at the stro customerID the s	e ries	Customers CustomerID CustomerID CustomerName RegionCode Country Employees Employees Employees Easthow/hide columns rer LabelName Reprise Malimba Records BSC Music The Classics Label (Laun sub) Music For Little People Metropolis Records TCB Music Winter & Winter Winter & Winter Winter & Winter Columbia River Entertainment Big Blue Dog DOMO Records DOMO Records DOMO Records	nove table rename tabl CategoryName POP RELAXATION NEW AGE CELLO ALT ROCK CHILDREN CHILDREN ALT ROCK JAZZ JAZZ JAZZ INDIA CHILDREN NEW AGE NEW AGE EASY LISTENING	e change t UnitPric 5.39 11.26 13.2 15.19 11.19 4.72 10.12 13.2 13.2 15.99 15.99 5.59 6.99 15.19 15.19 15.19	







Exercise:

1. Apply the OLAP operations for the banking application.



Signature of the Faculty

Information Technology

Page 45

Experiment 5: Implementation of Apriori Algorithm

Description:

The Apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. It uses a "bottom-up" approach, where frequent subsets are extended one at a time (a step known as candidate generation, and groups of candidates are tested against the data).

* <u>Problem:</u>

TID	ITEMS
100	1,3,4
200	2,3,5
300	1,2,3,5
400	2,5

To find frequent item sets for above transaction with a minimum support of 2 having confidence measure of 70% (i.e, 0.7).

Procedure:

<u>Step 1:</u>

Count the number of transactions in which each item occurs

TID	ITEMS
1	2
2	3
3	3
4	1
5	3

<u>Step 2:</u>

Eliminate all those occurrences that have transaction numbers less than the minimum support (2 in this case).

ITEM	NO. OF TRANSACTIONS
1	2
2	3
3	3
5	3

This is the single items that are bought frequently. Now let's say we want to find a pair of items that are bought frequently. We continue from the above table (Table in step 2).

<u>Step 3:</u>

We start making pairs from the first item like 1,2;1,3;1,5 and then from second item like 2,3;2,5. We do not perform 2,1 because we already did 1,2 when we were making pairs with 1 and buying 1 and 2 together is same as buying 2 and 1 together. After making all the pairs we get,

ITEM PAIRS	
1,2	
1,3	
1,5	
2,3	
2,5	
3,5	

Step 4:

Now, we count how many times each pair is bought together.

ITEM PAIRS	NO.OF TRANSACTIONS		
1,2	1		
1,3	2		
1,5	1		
2,3	2		
2,5	3		
3,5	2		

Step 5:

Again remove all item pairs having number of transactions less than 2.

ITEM PAIRS	NO. OF TRANSACTIONS
1,3	2
2,3	2
2,5	3
3,5	2

These pair of items is bought frequently together. Now, let's say we want to find a set of three

items that are bought together. We use above table (of step 5) and make a set of three items.

<u>Step 6:</u>

To make the set of three items we need one more rule (It is termed as self-join), it simply means, from item pairs in above table, we find two pairs with the same first numeric, so, we get (2,3) and (2,5), which gives (2,3,5). Then we find how many times (2, 3, 5) are bought together in the original table and we get the following

ITEM	NO. OF
SET	TRANSACTIONS
(2,3,5)	2

Thus, the set of three items that are bought together from this data are (2, 3, 5).

Confidence:

We can take our frequent item set knowledge even further, by finding association rules using the frequent item set. In simple words, we know (2, 3, 5) are bought together frequently, but what is the association between them. To do this, we create a list of all subsets of frequently bought items (2, 3, 5) in our case we get following subsets:

- {2}
- {3}
- {5}
- {2,3}
- {3,5}
- {2,5}

Now, we find association among all the subsets.

 $\{2\} => \{3,5\}:(If_{,,2}^{\circ\circ})$ is bought, what "s the probability that "3" and "5" would be bought in same transaction)

```
Confidence = P (3 \Box 5 \Box 2)/P(2) =2/3 =67%

{3}=>{2,5}= P (3 \Box 5 \Box 2)/P(3)=2/3=67%

{2,3}=>{5}= P (3 \Box 5 \Box 2)/P(2 \Box 3)=2/2=100%

{3,5}=>{2}= P (3 \Box 5 \Box 2)/P(3 \Box 5)=2/2=100%

{2,5}=>{3}= P (3 \Box 5 \Box 2)/P(2 \Box 5)=2/3=67%

Also, considering the remaining 2-items sets, we would get the following associations-

{1}=>{3}=P(1 \Box 3)/P(1)=2/2=100%

{3}=>{1}=P(1 \Box 3)/P(3)=2/3=67%

{2}=>{3}=P(3 \Box 2)/P(2)=2/3=67%

{3}=>{2}=P(3 \Box 2)/P(3)=2/3=67%

{2}=>{5}=P(2 \Box 5)/P(2)=3/3=100%

{5}=>{2}=P(2 \Box 5)/P(5)=3/3=100%
```

 $\{3\} => \{5\} = P(3 \Box 5)/P(3) = 2/3 = 67\%$

 $\{5\} => \{3\} = P(3 \Box 5)/P(5) = 2?3 = 67\%$

Eliminate all those having confidence less than 70%. Hence, the rules would be $-\{2,3\}=>\{5\}, \{3,5\}=>\{2\}, \{1\}=>\{3\}, \{2\}=>\{5\}, \{5\}=>\{2\}.$

Now these manual results should be checked with the rules generated in WEKA.

	Ciippoard	a ia		Font		ы	
	A1	•	6	<i>f</i> * 11			
							_
	Α	В	С	D	E	F	
1	11	12	13	14	15		
2	t		t	t			
3		t	t		t		
4	t	t	t		t		
5		t			t		
6							
7							
8							
9							

So first create a csv file for the above problem, the csv file for the above problem will look like the rows and columns in the above figure. This file is written in excel sheet.

Procedure for running the rules in weka:

Step 1:

Open weka explorer and open the file and then select all the item sets. The figure gives a better understanding of how to do that.



<u>Step 2:</u>

Now select the association tab and then choose apriori algorithm by setting the minimum support and confidence as shown in the figure

🕝 weka.gui.GenericObjectE	ditor	x	
weka.associations.Apriori			
About		-	
Class implementing a	n Apriori-type algorithm.	More	
	C	apabilities	
	· · · · · · · · · · · · · · · · · · ·		
car	False	•	
classIndex	-1		
Glassinger			
delta	0.05		
doNotCheckCapabilities	False	•	
lowerBoundMinSupport	0.4		
metricType	Confidence	•	
minMetric	0.7		
numRules	10		
	(False		
outputItemSets	False	•	
removeAllMissingCols	False	•	
significanceLevel	-1.0		
treatZeroAsMissing	False	v	
upperBoundMinSupport	1.0	The second secon	
•			
Open	Save OK	Cancel	
	Save	Cancer	

<u>Step 3:</u>

Now run the apriori algorithm with the set values of minimum support and the confidence. After running the weka generates the association rules and the respective confidence with minimum support as shown in the figure.

The above csv file has generated 5 rules as shown in the figure:

Associator	
and the second sec	
Choose Apricri -N10-T 0	-C07-D0.85-U10-MII.4-E-1.0-C-1
Associa	for output
Start Stop	
Result list(right-clic	Aus information ===
THE REAL PROPERTY AND	
17:05:47 - Apriori Scher	mer: weka.awsocistions.Apriori -W 10 -F 3 -C 0.7 -D 0.05 -D 1.0 -M 0.4 -S -1.0 -C -1 tion: Booki
	sides: 4
1.000000	Lbotes: S
	Il
	п
	11
	14
	15
)	Associator model (full training set) ===
Aprii	
001.40	
Minis	num support: 0.4 (2 irstances)
1.30074	mum metric <confidence>: 0.7</confidence>
Raniby	er of cycles performed: 12
Gener	rated sets of large itemsets:
Sibe	of set of large itemsets L[1]: 4
Size	of set of large itemzets L(2): 4
Sipe	of set of large itemzets L(3): 1

15	
=== Associator model (full training set) ===	
Apriori	
Minimum support: 0.4 (2 instances)	
Minimum metric <confidence>: 0.7</confidence>	
Number of cycles performed: 12	
Generated sets of large itemsets:	
Size of set of large itemsets L(1): 4	
Size of set of large itemsets $L(2)$: 4	
Size of set of large itemsets L(3): 1	
Best rules found:	
1. I5=t 3 ==> I2=t 3 <conf:(1)> lift:(1.33) lev:(0.19) [0] conv:(0.75)</conf:(1)>	
2. I2=t 3 ==> I5=t 3 <conf:(1)> lift:(1.33) lev:(0.19) [0] conv:(0.75)</conf:(1)>	
3. I1=t 2 ==> I3=t 2 <conf:(1)> lift:(1.33) lev:(0.13) [0] conv:(0.5)</conf:(1)>	
4. I3=t I5=t 2 ==> I2=t 2 <conf: (1)=""> lift: (1.33) lev: (0.13) [0] conv: (0.5) 5. I2=t I3=t 2 ==> I5=t 2 <conf: (1)=""> lift: (1.33) lev: (0.13) [0] conv: (0.5)</conf:></conf:>	

Conclusion:

As we have seen the total rules generated by us manually and by the weka are matching, hence the rules generated are 5.

Exercise:

1. Apply the Apriori algorithm on Airport noise monitoring dataset discriminating between patients with parkinsons and neurological diseases using voice recording dataset. [https://archive.ics.uci.edu/ml/machine-learning-databases/00000/ refer this link for datasets]



Signature of the Faculty

Experiment 6: Implementation of FP- Growth Algorithm

(6a) Aim: To generate association rules using FP Growth Algorithm

PROBLEM:

To find all frequent item sets in following dataset using FP-growth algorithm. Minimum support=2 and confidence =70%

TID	ITEMS
100	1,3,4
200	2,3,5
300	1,2,3,5
400	2,5

Solution:

Similar to Apriori Algorithm, find the frequency of occurrences of all each item in dataset and then prioritize the items according to its descending order of its frequency of occurrence. Eliminating those occurrences with the value less than minimum support and assigning the priorities, we obtain the following table.

ITEM	NO. OF TRANSACTIONS	PRIORITY
1	2	4
2	3	1
3	3	2
5	3	3

Re-arranging the original table, we obtain

TID	ITEMS
100	1,3
200	2,3,5
300	2,3,5,1
400	2,5

Construction of tree:

Note that all FP trees have ",null" node as the root node. So, draw the root node first and attach the items of the row 1 one by one respectively and write their occurrences in front of it. The tree is further expanded by adding nodes according to the prefixes (count) formed and by further incrementing the occurrences every time they occur and hence the tree is built.

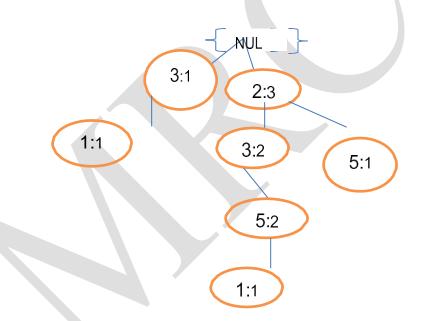
Prefixes:

- 1->3:1 2,3,5:1
- **5**->2,3:2 2:1
- 3->2:2

Frequent item sets:

- 1-> 3:2 /*2 and 5 are eliminated because they "re less than minimum support, and the occurrence of 3 is obtained by adding the occurrences in both the instances*/
- Similarly, 5->2,3:2 ; 2:3;3:2
- 3->2 :2

Therefore, the frequent item sets are $\{3,1\}$, $\{2,3,5\}$, $\{2,5\}$, $\{2,3\}$, $\{3,5\}$ The tree is constructed as below:



Generating the association rules for the following tree and calculating the confidence measures we get-

- {3}=>{1}=2/3=67%
- {1}=>{3}=2/2=100%
- {2}=>{3,5}=2/3=67%

- {2,5}=>{3}=2/3=67%
- $\{3,5\} => \{2\} = 2/2 = 100\%$
- {2,3}=>{5}=2/2=100%
- {3}=>{2,5}=2/3=67%
- {5}=>{2,3}=2/3=67%
- {2}=>{5}=3/3=100%
- $\{5\} => \{2\} = 3/3 = 100\%$
- {2}=>{3}=2/3=67%
- {3}=>{2}=2/3=67%

Thus eliminating all the sets having confidence less than 70%, we obtain the following conclusions:

$$\{1\} => \{3\}, \{3,5\} => \{2\}, \{2,3\} => \{5\}, \{2\} => \{5\}, \{5\} => \{2\}.$$

As we see there are 5 rules that are being generated manually and these are to be checked against the results in WEKA. Inorder to check the results in the tool we need to follow the similar procedure like

Apriori.

				FORT		
	Cippoard	1 is		Font		La l
	A1	-		<i>f</i> * 11		
	А	В	С	D	E	F
1	11	12	13	14	15	
2	t		t	t		
3		t	t		t	
4	t	t	t		t	
5		t			t	
6						
7						
8						
9						

Sofirst create a csvfile for the above problem, the csv file for the above problem will look like the rows and columns in the above figure. This file is written in excel sheet.

Procedure for running the rules in weka:

<u>Step 1:</u>

Open weka explorer and open the file and then select all the item sets. The figure gives a better understanding of how to do that.

Current relation	Selected attr	ribute		
Relation: Book1 Attributes: 5 Instances: 4 Sum of weights: 4	Name: Missing: 1		Distinct: 1	Type: Nominal Unique: 0 (0%)
Attributes	No.	Label	Count	Weight
All None Invert Pattern			2	2.0
1 √ 11 2 √ 12 3 √ 13 4 √ 14	Class: 15 (No	om)		Visualize All
5 🗹 15	2			
5 🗹 15	2			
5 🗹 15				Log x0

<u>Step 2:</u>

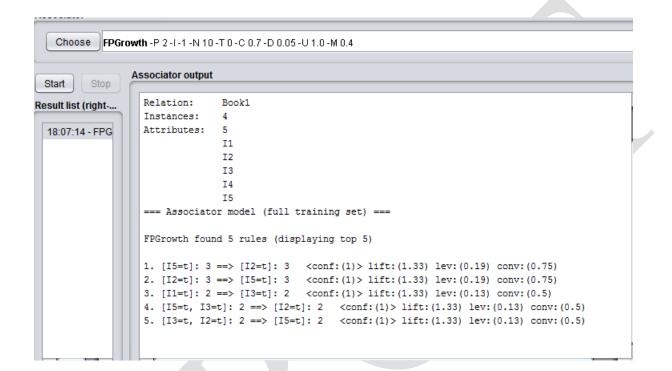
Now select the association tab and then choose FP growth algorithm by setting the minimum support and confidence as shown in the figure.

ka.associations.FPGrowth		
Jour		
		ore
without candidate generati	on. Capa	bilities
1.		
delta	0.05	
doNotCheckCapabilities	False	•
dAllRulesForSupportLevel	False	-
IowerBoundMinSupport	0.1	
maxNumberOfItems	-1	
metricType	Confidence	-
minMetric	0.9	
numRulesToFind	10	
positiveIndex	2	
rulesMustContain		
transactionsMustContain		
upperBoundMinSupport	1.0	
useORForMustContainList	False	-
useorcronwusicontainList	[Fdise	

Step 3:

Now run the FP Growth algorithm with the set values of minimum support and the confidence. After running the weka generates the association rules and the respective confidence with minimum support as shown in the figure.

The above csv file has generated 5 rules as shown in the figure:



Conclusion:

As we have seen the total rules generated by us manually and by the weka are matching, hence the rules generated are 5.

Exercise

1. Apply FP-Growth algorithm on Blood Transfusion Service Center data set



Signature of the Faculty

Information Technology

Page 64

Experiment 7: Decision Tree Induction

Aim: Generate a Decision Tree by using J48 algorithm.

DESCRIPTION:

Decision tree learning is one of the most widely used and practical methods for inductive inference over supervised data. It represents a procedure for classifying categorical database on their attributes. This representation of acquired knowledge in tree form is intuitive and easy to assimilate by humans.

ILLUSTRATION:

Build a decision tree for the following data

AGE	INCOM E	STUDENT	CREDIT_RATING	BUYS_COMPUTER		
Youth	High	No	Fair	No		
Youth	High	No	Excellent	No		
Middle aged	Middle aged High No		No Fair			
Senior	Medium	No	Fair	Yes		
Senior	Low	Yes	Fair	Yes		
Senior	Low	Yes	Excellent	No		
Middle aged	Medium	Yes	Excellent	Yes		
Youth	Low	No	Fair	No		
Youth	Medium	Yes	Fair	Yes		
Senior	Medium	Yes	Fair	Yes		
Youth	Medium	Yes	Excellent	Yes		
Middle aged	Middle aged Medium No		Excellent	Yes		
Middle aged	High	Yes	Fair	Yes		
Senior	Medium	No	Excellent	No		

The entropy is a measure of the uncertainty associated with a random variable. As uncertainty increases, so does entropy, values range from [0-1] to present the entropy of information

Entropy (D) = $\sum_{j=1}^{c} -p \log_2 p$

Information gain is used as an attribute selection measure; pick the attribute having the highest information gain, the gain is calculated by:

Gain (D, A) = Entropy (D) - $\sum_{j=1}^{c} (|D_j|/|D|) Entropy(D)$

Where, D: A given data partition A: Attribute

V: Suppose we were partition the tuples in D on some attribute A having v distinct values D is split into v partition or subsets, (D1, D2...., Dj), where Dj contains those tuples in D that have outcome Aj of A.

Class P: buys_computer="yes" Class N: buys_computer="no"

Entropy (D) = $-9/14\log (9/14)-5/15\log (5/14) = 0.940$ Compute the expected information requirement for each attribute start with the attribute age Gain (age, D)

= Entropy (D) - $\sum_{youth,middle-agedsenior} \left(\frac{Sv}{s}\right) Entropy(Sv)$

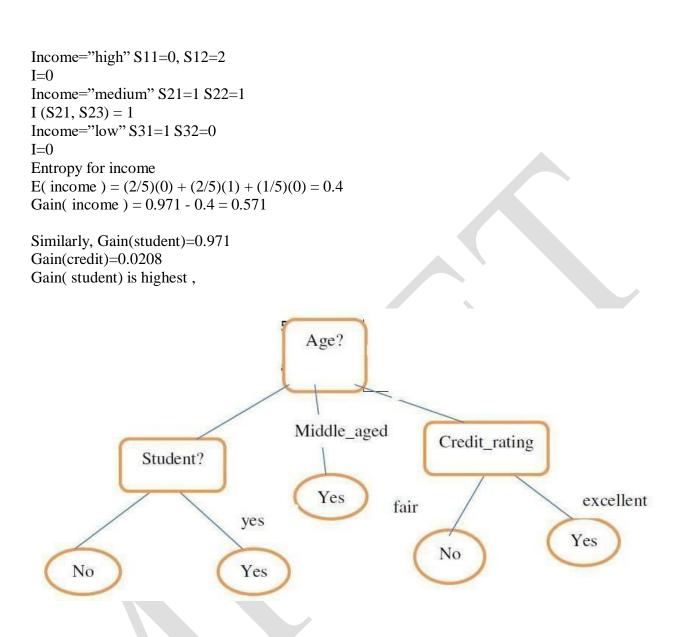
```
= Entropy (D) - 5/14Entropy(Syouth)-4/14Entropy(Smiddle-aged)-5/14Entropy(Ssenior)
= 0.940-0.694
=0.246
```

Similarly, for other attributes, Gain (Income, D) =0.029Gain (Student, D) = 0.151Gain (credit_rating, D) = 0.048

Income	Student	Credit_rating	Class
High	No	Fair	No
High	No	Excellent	No
Mediun	n No	Fair	No
Low	Yes	Fair	Yes
mediun	n Yes	excellent	yes

Now, calculating information gain for subtable (age<=30)

I The attribute age has the highest information gain and therefore becomes the splitting * attribute at the root node of the decision tree. Branches are grown for each outcome of age. These tuples are shown partitioned accordingly.



A decision tree for the concept buys_computer, indicating whether a customer at All Electronics is likely to purchase a computer. Each internal (non-leaf) node represents a test on an attribute. Each leaf node represents a class (either buys computer="yes" or buys computer="no".

first create a csv file for the above problem, the csv file for the above problem will look like the rows and columns in the above figure. This file is written in excel sheet.

	Clipboard	d G		Font		Fai
	A1		. (*	<i>f</i> ∗ age		
	А	В	С	D	E	F
1	age	income	student	credit_rat	buys_com	puter
2	<=30	high	no	fair	no	
3	<=30	high	no	excellent	no	
4	3140	high	no	fair	yes	
5	>40	medium	no	fair	yes	
6	>40	low	yes	fair	yes	
7	>40	low	yes	excellent	no	
8	3140	low	yes	excellent	yes	
9	<=30	medium	no	fair	no	
10	<=30	low	yes	fair	yes	
11	>40	medium	yes	fair	yes	
12	<=30	medium	yes	excellent	yes	
13	3140	medium	no	excellent	yes	
14	3140	high	yes	fair	yes	
15	>40	medium	no	excellent	no	
16						

Procedure for running the rules in weka:

<u>Step 1:</u>

Open weka explorer and open the file and then select all the item sets. The figure gives a better understanding of how to do that.



Step2:

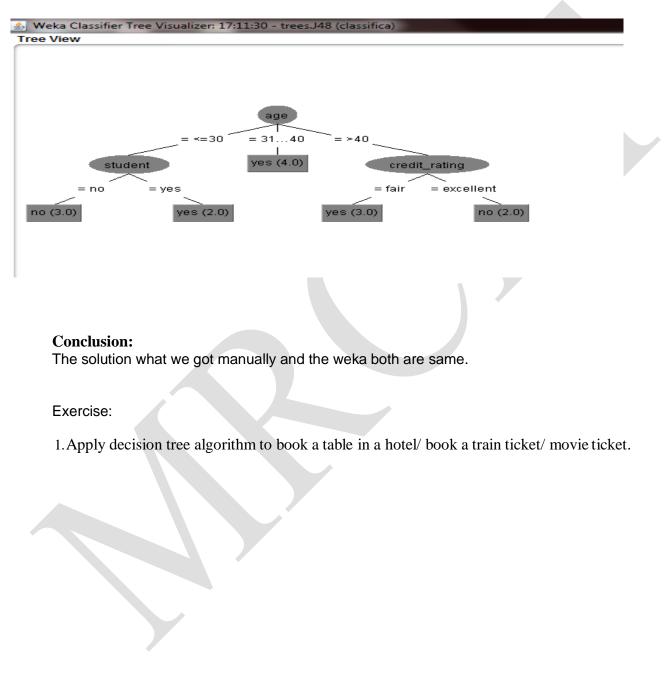
Now select the classify tab in the tool and click on start button and then we can see the result of the problem as below

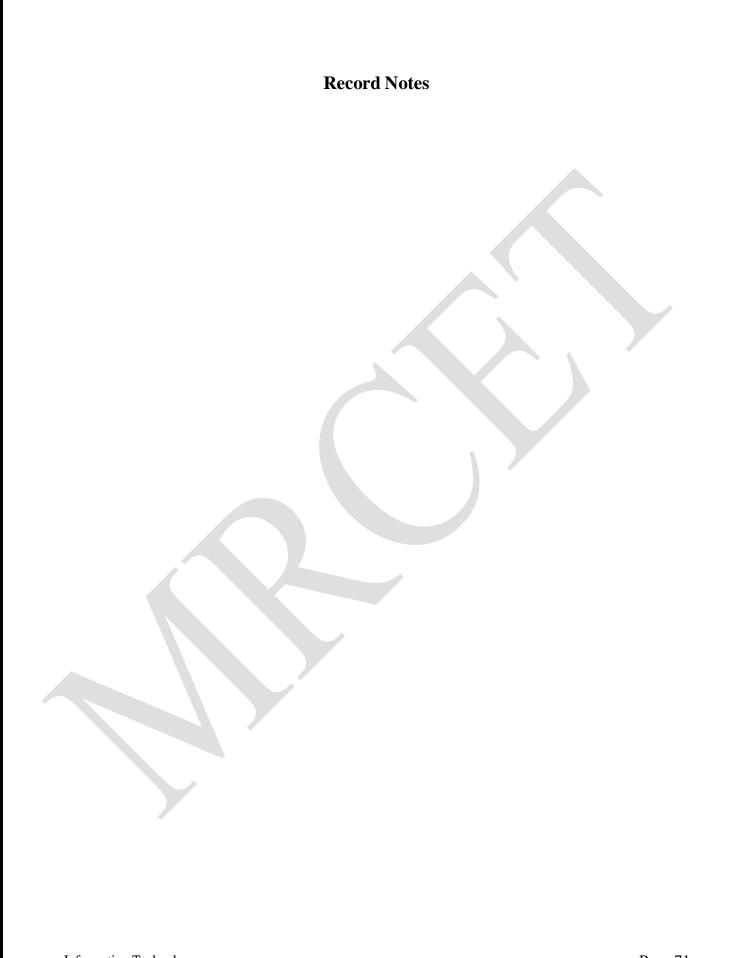
est options	Classifier output									
Use training set Supplied test set Set Cross-validation Folds 10	=== Stratified	Time taken to build model: 0.02 s === Stratified cross-validation = === Summary ===								
O Percentage split % 66	Correctly Class	Correctly Classified Instances				50	5			
More options	Incorrectly Cla			7		50	40 40			
more options	Kappa statistic			-0.04	26		~			
	Mean absolute e			0.41	.67					
Nom) buys_computer	Root mean squar	ed error		0.59	84					
ioni, odjo_compator	Relative absolu	te error		87.5	\$					
Start Stop	Root relative s	quared err	or	121.2987 %						
	Total Number of	Instances	1	14						
sult list (right-click for options)	Tables and the development for the									
17:11:30 - trees.J48	=== Detailed Ac	curacy By	Class ===							
17.11.50 - 1005.546		TP Rate	FP Rate	Precision	Recall.	F-Measure	MCC	ROC Area	PRC Area	Class
		0.400	0.444	0.333	0.400	0.364	-0.043	0.633	0.457	no
		0.556	0.600	0.625	0.556	0.588	-0.043	0.633	0.758	yes
	Weighted Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.633	0.650	0001000
	=== Confusion M	atrix ===								
	a b < clas	sified as								
	23 a = no									
	4 5 b = yes									

Step3:

Check the main result which we got manually and the result in weka by right clicking on the result and visualizing the tree.

The visualized tree in weka is as shown below:





Experiment 8: Calculating information gain measures.

Information gain (IG) measures how much "information" a feature gives us about the class. -Features that perfectly partition should give maximal information. – Unrelated features should give no information. It measures the reduction in entropy. CfsSubsetEval aims to identify a subset of attributes that are highly correlated with the target while not being strongly correlated with one another. It searches through the space of possible attribute subsets for the "best" one using the BestFirst search method by default, although other methods can be chosen. To use the wrapper method rather than a filter method, such as CfsSubsetEval, first select WrapperSubsetEval and then configure it by choosing a learning algorithm to apply and setting the number of crossvalidation folds to use when evaluating it on each attribute subset.

Steps:

- Open WEKA Tool.
- Click on WEKA Explorer.
- Click on Preprocessing tab button.
- Click on open file button.
- Select and Click on data option button. .
- Choose a data set and open file. •
- Click on select attribute tab and Choose attribute evaluator, search method algorithm

Weka Explorer		
	Select attributes Visualize	
Attribute Evaluator		
Choose InfoGainAttributeEval		
Search Method		
Choose Ranker -T -1.79769313486231	7E308 -N -1	
Attribute Selection Mode	Attribute selection output	
Use full training set Cross-validation Folds 10 Seed 1 (Nom) class Start Stop Result list (right-click for options) 12:06:47 - BestFirst + CfsSubsetEval 12:07:06 - GreedyStepwise + CfsSubsetE 12:07:57 - Ranker + InfoGainAttributEval 12:08:03 - Ranker + InfoGainAttributEval	<pre>=== Run information === Evaluator: weka.attributeSelection.InfoGainAttributeEval Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1 Relation: iris Instances: 10 Attributes: 5 sepallength petallength petallength class Evaluation mode: evaluate on all training data === Attribute Selection on all input data ===</pre>	
Status OK	Search Method: Attribute ranking. Attribute Evaluator (supervised, Class (nominal): 5 class): Information Gain Ranking Filter	Log

CK OH STAIT DUITON.

Preprocess Classify Cluster Assoc	ate Select attributes Visualize	
Attribute Evaluator		
Choose InfoGainAttributeEval		
Search Method		
Choose Ranker -T -1.797693134862	3157E308 -N -1	
Attribute Selection Mode	Attribute selection output	
Ouse full training set Cross-validation Folds Seed	petallength petalwidth class Evaluation mode: evaluate on all training data	<u> </u>
(Nom) class Start Stop Result list (right-click for options)	=== Attribute Selection on all input data === Search Method: Attribute sanking.	
12:06:47 - BestFirst + CfsSubsetEval 12:07:06 - GreedyStepwise + CfsSubsetE 12:07:57 - Ranker + InfoGainAttributeEval	Attribute Evaluator (supervised, Class (nominal): 5 class):	
12:08:03 - Ranker + InfoGainAttributeEval	kaned attributes: 0 4 petalwidth 0 3 petallength 0 2 sepalwidth 0 1 sepallength	
•	Selected attributes: 4,3,2,1 : 4	, T
Status		

Exercise

Calculate the information gain on weather data set(for each attributes separately).



Experiment 9: Classification of data using Bayesian approach

AIM: To apply naïve bayes classifier on a given data set.

Description:

In machine learning, Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes" Theorem with strong (naïve) independence assumptions between the features

Example:

AGE	INCOME	STUDENT	CREDIT_RATING	BUYS_COMPUTER
<30 High		No	Fair	No
<30	<30 High N		Excellent	No
31-40	High	No	Fair	Yes
>40	Mediu m	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31-40	Mediu m	Yes	Excellent	Yes
<=30	Low	No	Fair	No
<=30	Mediu m	Yes	Fair	Yes
>40	Mediu m	Yes	Fair	Yes
<30	Mediu m	Yes	Excellent	Yes
31-40	Mediu m	No	Excellent	Yes
31-40	High	Yes	Fair	Yes
>40	Mediu m	No	Excellent	No

CLASS:

C1:buys_com puter = 'yes' C2:buys_com puter='no' DATA TO BECLASSIFIED

:

X=(age<=30, income=Medium, Student=Yes, credit_rating=Fair)

• P(C1): P(buys_computer="yes")= 9/14 = 0.643

P (buys_computer="no") =5/14=0.357

- Compute P(X/C1) and p(x/c2) weget:
 - 1. P(age="<=30" |buys_computer="yes")=2/9
 - 2. P(age="<=30"|buys_computer="no")=3/5
 - 3. P(income="medium"|buys_computer="yes")=4/9
 - 4. P(income="medium"|buys_computer="no")=2/5
 - 5. P(student="yes"|buys_computer="yes")=6/9
 - 6. P(student="yes" |buys_computer="no")=1/5=0.2
 - 7. P(credit_rating="fair"|buys_computer="yes")=6/9
 - 8. P(credit_rating="fair" |buys_computer="no")=2/5
 - X=(age<=30, income=medium, student=yes, credit_rating=fair) P(X/C1): P (X/buys_computer="yes")=2/9*4/9*6/9*6/9= 32/1134 P(X/C2):P(X/buys_computer="no")=3/5*2/5*1 /5*2/5=12/125

P(C1/X)=P(X/C1)*P(C1)

P(X/buys_computer="yes")*P(buys_computer="yes")=(32/1134)*(9/14)=0.019

P(C2/X)=p(x/c2)*p(c2)

P (X/buys_computer="no")*P(buys_computer="no")=(12/125)*(5/14)=0.007

Therefore, conclusion is that the given data belongs to C1 since P(C1/X)>P(C2/X)

Checking the result in the WEKA tool:

In order to check the result in the tool we need to follow a procedure.

Step:1

Create a csv file with the above table considered in the example. the arfffile will look as shown below:

🔲 store.arff - Notepad
File Edit Format View Help
@relation store
<pre>@attribute age {young, middle-aged, old} @attribute income {high,medium,low} @attribute student {yes,no} @attribute credit-rating {fair,excellent} @attribute buys-computer {yes,no}</pre>
@data young,high,no,fair,no young,high,no,excellent,no middle-aged,high,no,fair,yes old,medium,no,fair,yes old,low,yes,fair,yes old,low,yes,excellent,no middle-aged,low,yes,excellent,yes young,medium,no,fair,no young,low,yes,fair,yes old,medium,yes,fair,yes old,medium,yes,fair,yes middle-aged,medium,no,excellent,yes middle-aged,high,yes,fair,yes old,medium,no,excellent,no

Step2:

Now open weka explorer and then select all the attributes in the table.

Preprocess Classify Cluster Associate Select attributes	Visualize				
Open file Open URL Open DB Ger	nerate		Undo	Edit	Save
Filter					
Choose None					Apr
Current relation	Selecte	d atl	tribute		
Relation: classifica Attributes: 5 Instances: 14 Sum of weights: 14			age 0 (0%)	Distinct: 3	Type: Nominal Unique: 0 (0%)
Attributes	No.		Label	Count	Weight
	1	1	<=30	5	5.0
			3140	4	4.0
All None Invert Pattern	-	3	>40	5	5.0
No. Name					
1 🗹 age	Class	NUVS	_computer (No	m)	Visualiz
2 🗹 income	(0.000.	Jajo.		,	
3 🗹 student					
4 🗹 credit_rating	5				5
5 🗹 buys_computer			4	Ĩ.	
Remove					1.5

<u>Step 3:</u>

Select the classifier tab in the tool and choose baye"s folder and then naïve baye"s classifier to see the result as shown below.

ssifier output	on test se	·c ====							
lime taken to te	st model	on suppli	ed test set	: 0.01 s	conds				
=== Summary ===									
Correctly Classi	fied Inst	ances	0		0	ŧ			
incorrectly Clas	sified In	stances	1		100	\$			
(appa statistic			0						
Mean absolute er	ror		0.75	38					
loot mean square	d error		0.75	38					
Relative absolut	e error		120.61	.24 %					
Root relative so	quared err	or	120.61	.24 %					
otal Number of	Instances		1						
== Detailed Acc	uracy By	Class ===							
	0.000	1.000	Precision 0.000 0.000	0.000		0.000	ROC Area ? ?	PRC Area ? 1.000	Class yes no
Weighted Avg.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	
	trix ===								
=== Confusion Ma									

Exercise

1. Classify data (lung cancer/ diabetes /liver disorder) using Bayesian approach .



Experiment 10: Implementation of K-means algorithm

DESCRIPTION:

K-means algorithm aims to partition n observations into "k clusters" in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in partitioning of the data into Voronoi cells.

ILLUSTRATION:

As a simple illustration of a k-means algorithm, consider the following data set consisting of the scores of two variables on each of the five variables.

I	X1	X2
А		1
В	1	0
С	0	2
D	2	4
E	3	5

This data set is to be grouped into two clusters: As a first step in finding a sensible partition, let the A & C values of the two individuals furthest apart (using the Euclidean distance measure), define the initial cluster means, giving:

Cluster	Individual	Mean Vector(Centroid)
Cluster 1	А	(1,1)
Cluster2	С	(0,2)

The remaining individuals are now examined in sequence and allocated to the cluster to which they are closest, in terms of Euclidean distance to the cluster mean. The mean vector is recalculated each time a new member is added. This leads to the following series of steps:

	A	С
А	0	1.4
В	1	2.5
С	1.4	0
D	3.2	2.82
Е	4.5	4.2

Initial partitions have changed, and the two clusters at this stage having the following characteristics.

	Individual	Mean vector(Centroid)
Cluster 1	A,B	(1,0.5)
Cluster 2	C,D,E	(1.7,3.7)

But we cannot yet be sure that each individual has been assigned to the right cluster. So, we compare each individual's distance to its own cluster mean and to that of the opposite cluster. And, we find:

Ι	А	С
А	0.5	2.7
В	0.5	3.7
С	1.8	2.4
D	3.6	0.5
E	4.9	1.9

The individuals C is now relocated to Cluster 1 due to its less mean distance with the centroid points. Thus, its relocated to cluster 1 resulting in the new partition

	Individual	Mean vector(Centroid)
Cluster 1	A,B,C	(0.7,1)
Cluster 2	D,E	(2.5,4.5)

The iterative relocation would now continue from this new partition until no more relocation occurs. However, in this example each individual is now nearer its own cluster mean than that of the other cluster and the iteration stops, choosing the latest partitioning as the final cluster solution. Also, it is possible that the k-means algorithm won't find a final solution. In this case, it would be a better idea to consider stopping the algorithm after a pre-chosen maximum number of iterations.

Checking the solution in weka:

In order to check the result in the tool we need to follow a procedure.

<u>Step 1:</u>

Create a csv file with the above table considered in the example. the csv file will look as shown below:

Clipboard				Fant	
	Clipboard	1 6		Font	
A1		,	- (*	<i>f</i> ∗ i	
	1	1	1		
	Α	В	С	D	
1	i	x1	x2		
2	Α	1	. 1		
3	В	1	. 0		
4	С	C	2		
5	D	2	. 4		
6	E	3	5		
7					
		1			

Step 2:

Now open weka explorer and then select all the attributes in the table.

<u></u>	None					Apply
urrent relat	ion	Selected at	tribute			
	Relation: menas Attributes: 3 Instances: 5 Sum of weights: 5		Name: i Missing: 0 (0%) Distinct: 5		Type: Nominal Unique: 5 (100%)	
ttributes		No.	Label	Count	Weight	
Martin Martin De De		1	A	1	1.0	
C		234	В	1	1.0	
All	None Invert Pattern	3	С	1	1.0	
		4	D	1	1.0	
No.	Name					
1	√ i	Class: x2 (N	um)		V V	/isualize Al
2	✓ x1 ✓ x2		5,			no danzo ra
3						
		23				
		1	1	- ³		1
		1	1		1	1
		1	1		1	1
					1	1
	Remove		1		1 6	1
		1		- 1		1
			. 1			1
tatus						1 • • • • ×

<u>Step 3:</u>

Select the cluster tab in the tool and choose normal k-means technique to see

the result as shown below.

=== Run info:	rmation ===	
Scheme:	weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -	-A "
Relation:	menas	
Instances:	5	
Attributes:	3	
	1	
	x1	
	x2	
Test mode:	evaluate on training data	
=== Clusterin	ng model (full training set) ===	
kMeans		
xiicuiis		
Number of its		
Within clust	er sum of squared errors: 3.22962962963	
	ting points (random):	
Initial star		
	.2,4	
Initial star Cluster 0: D Cluster 1: B		

Final cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(5.0)	(2.0)	(3.0)
i	A	D	A
x 1	1.4	2.5	0.6667
x2	2.4	4.5	1

Time taken to build model (full training data) : 0 seconds === Model and evaluation on training set === Clustered Instances 0 2 (40%) 1 3 (60%)

Exercise

1. Implement of K-means clustering using crime dataset.



Information Technology

Page 93

Case Study

Create Placement.arff file to identify the students who are eligible for placements using KNN

Information Technology

Page 97